

Open Research Online

The Open University's repository of research publications and other research outputs

A new tool for the validation of dynamic simulation models

Thesis

How to cite:

Martin, Christopher John (1995). A new tool for the validation of dynamic simulation models. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 1995 The Author



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.0000e0d0>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

**A NEW TOOL FOR THE VALIDATION
OF DYNAMIC SIMULATION MODELS**

A thesis submitted to the Open University for the
degree of Doctor of Philosophy in the discipline of
Energy and Environmental Research

Christopher John Martin MA (Cantab), MSE

June 1995

BEST COPY

AVAILABLE

Variable print quality

ABSTRACT

A new method for determining the causes of discrepancies between dynamic simulation models and measured data is described. Whilst principally aimed at the validation of building thermal simulation codes, the method may find application in the validation of simulation models from other disciplines. The new method relies on generating a model of the discrepancies in terms of the variables driving the physical reality which produced the experimental data and the corresponding simulation. Inspection of the resulting error model allows the principal contributors to simulation error to be identified, its dynamic nature to be characterised, and the likely cause of the error to be identified.

The method is tested by generating a 'quasi-truth' dataset using the building thermal simulation code SERI-RES. A series of simulations is then prepared with perturbations to selected input parameters. The new tool is found to be capable of recovering details of the perturbations.

Finally, the power of the new technique is demonstrated in a series of comparisons between the predictions of the model SERI-RES and data collected in outdoor test rooms. These comparisons reveal that the principal source of error in the predictions is the treatment of the interactions between the heater and the air and fabric of the room. The new technique proves sufficiently sensitive to detect the changing structure of the prediction errors as the position of the heater within the room is changed.

ACKNOWLEDGEMENTS

This thesis is in part the product of many absorbing discussions and lively arguments with workers throughout the building simulation field. I would particularly like to thank Ben Bland, David Bloomfield, Joe Clarke, Pascal Dalicieux, Tim Dewson, Herbert Eppel, Martin Gough, Alan Guy, Geoff Hammond, Andy Irving, Alan Jones, Ron Judkoff, Kevin Lomas, Nacim Ramdani and Paul Strachan.

Special thanks are due to Martin Watson, for his exacting approach to our experimental work over the past ten years, and to my two long suffering supervisors, Quentin Jones and John Wiltshire.

However, the last word on validation, particularly the validation of building thermal simulation models, has to go to Bob Dylan:

'You either got faith, or you got unbelief.

There ain't no neutral ground.'

CONTENTS

	page
1 INTRODUCTION	
2 A REVIEW OF MODEL VALIDATION	
2.1 What is model validation ?	2-1
2.1.1 The distinction between verification and validation	2-3
2.1.2 The importance of context in model testing	2-4
2.1.3 Model validation and falsifiability	2-6
2.2 Approaches to model testing	2-7
2.2.1 Testing in the context of a three stage modelling process	2-10
2.2.2 Tools for model testing	2-11
2.3 Validation of building thermal simulation models	2-15
2.4 Tools for establishing model validity	2-28
2.4.1 Statistical tests for model validity	2-28
2.4.2 Output uncertainty criteria for validity	2-34
2.4.3 Statistical validity criteria versus output uncertainty methods	2-37
2.5 Tools for diagnosing the reasons for poor predictions	2-38
3 A NEW METHOD FOR IDENTIFYING SOURCES OF ERROR	
3.1 A linear time-invariant error model	3-1
3.1.1 State-space models	3-2
3.1.2 Auto-regressive/moving average (ARMA) models	3-4
3.1.3 Transfer function modelling	3-5
3.1.4 Frequency response modelling	3-7
3.1.5 Impulse response models	3-8
3.2 Selection of an error model representation	3-9
4 SOLUTION OF THE IMPULSE RESPONSE MODEL	
4.1 Single input case	4-1
4.1.1 Deconvolution using cross-covariance functions	4-6
4.2 Multiple input case	4-8
4.3 Propagation of mean values through the deconvolution process	4-11
4.4 Implementation	4-13

4.5	Verification of the solution process	4-15
4.6	Guidelines for the design of impulse response identification experiments	4-24

5 TESTING THE METHOD

5.1	Analytical trials with SERI-RES	5-1
5.2	Identification of sources of simulation error	5-6
5.2.1	Conductive heat loss effects	5-8
5.2.2	Solar radiation effects	5-12
5.2.3	Thermal capacity effects	5-14
5.3	Correlation and causality	5-14
5.4	Conclusions and guidelines for the design of validation experiments	5-15
5.5	The rôle of the new technique in the model testing process	5-17

6 VALIDATION USING REAL DATA

6.1	Choice of buildings for validation	6-1
6.1.1	The Energy Monitoring Company Outdoor Test Room facility	6-3
6.2	Operating buildings for model validation	6-4
6.3	Operating a simulation model for model validation	6-7
6.4	Details of the dataset	6-10
6.5	Carrying out simulations for comparison with the measured data	6-12
6.5.1	Meteorological data handling	6-12
6.5.2	Simulation scheduling	6-14
6.5.3	Modification to predicted floor surface quantities	6-17
6.6	Comparisons between simulation results and measured data	6-18
6.7	Detailed analysis of simulation error structure using the new method	6-20
6.7.1	Zone temperature predictions	6-21
6.7.2	Surface temperature predictions	6-21
6.7.3	Surface heat flux predictions	6-22
6.8	Analysis of predicted and measured responses to heater power	6-22
6.9	Conclusions of the empirical validation experiment	6-23

7 CONCLUSIONS

BIBLIOGRAPHY

FIGURES

- Figure 1.1 Testing and application of the new method
- Figure 4.1 Analytical test of impulse response model solution: Response to test input stream 1
- Figure 4.2 Analytical test of impulse response model solution: Response to test input stream 2
- Figure 4.3 Analytical test of impulse response model solution: Response to test input stream 3
- Figure 4.4 Analytical test of impulse response model solution: Response to test input stream 4
- Figure 4.5 Analytical test of impulse response model solution: Response to test input stream 5
- Figure 4.6 Analytical test of impulse response model solution: Variation of estimation error with number of points
- Figure 5.1 Effect of perturbing test room Glazing U-value: Zone temperature prediction
- Figure 5.2 Effect of perturbing test room Glazing U-value: Floor surface temperature prediction
- Figure 5.3 Effect of perturbing test room Floorslab density: Zone temperature prediction
- Figure 5.4 Effect of perturbing test room Floorslab density: Floor surface temperature prediction
- Figure 5.5 Sensitivity to error in Glazing U-value: Zone temperature prediction error responses
- Figure 5.6 Sensitivity to error in Glazing U-value: Floor temperature prediction error responses
- Figure 5.7 Sensitivity to error in Glazing U-value: Floor heat flux prediction error responses
- Figure 5.8 Sensitivity to error in Glazing U-value: Back wall temperature prediction error responses
- Figure 5.9 Sensitivity to error in Glazing U-value: Back wall heat flux prediction error responses

Figure 5.10	Sensitivity to errors in Glazing U-value: Contributions of driving forces to zone temperature error
Figure 5.11	Sensitivity to errors in Glazing U-value: Reconstituted and observed zone temperature errors
Figure 5.12	Sensitivity to error in Glazing U-value: Relative contributions of driving forces to prediction errors
Figure 5.13	Changes in surface heat flux when glazing U-value is increased
Figure 5.14	Sensitivity to error in Styrofoam conductivity: Zone temperature prediction error responses
Figure 5.15	Sensitivity to error in Styrofoam conductivity: Floor temperature prediction error responses
Figure 5.16	Sensitivity to error in Styrofoam conductivity: Floor heat flux prediction error responses
Figure 5.17	Sensitivity to error in Styrofoam conductivity: Back wall temperature prediction error responses
Figure 5.18	Sensitivity to error in Styrofoam conductivity: Back wall heat flux prediction error responses
Figure 5.19	Sensitivity to error in Styrofoam conductivity: Relative contributions of driving forces to prediction errors
Figure 5.20	Changes in surface heat flux when Floor U-value is increased
Figure 5.21	Sensitivity to error in Solar-loss: Zone temperature prediction error responses
Figure 5.22	Sensitivity to error in Solar-loss: Floor temperature prediction error responses
Figure 5.23	Sensitivity to error in Solar-loss: Floor heat flux prediction error responses
Figure 5.24	Sensitivity to error in Solar-loss: Back wall temperature prediction error responses
Figure 5.25	Sensitivity to error in Solar-loss: Back wall heat flux prediction error responses
Figure 5.26	Sensitivity to error in Solar-loss: Relative contributions of driving forces to prediction errors
Figure 5.27	Sensitivity to error in Solar distribution: Zone temperature prediction error responses
Figure 5.28	Sensitivity to error in Solar distribution: Floor temperature prediction error responses

Figure 5.29	Sensitivity to error in Solar distribution: Floor heat flux prediction error responses
Figure 5.30	Sensitivity to error in Solar distribution: Back wall temperature prediction error responses
Figure 5.31	Sensitivity to error in Solar distribution: Back wall heat flux prediction error responses
Figure 5.32	Sensitivity to error in Solar distribution: Relative contributions of driving forces to prediction errors
Figure 5.33	Sensitivity to error in Floorslab density: Zone temperature prediction error responses
Figure 5.34	Sensitivity to error in Floorslab density: Floor temperature prediction error responses
Figure 5.35	Sensitivity to error in Floorslab density: Floor heat flux prediction error responses
Figure 5.36	Sensitivity to error in Floorslab density: Back wall temperature prediction error responses
Figure 5.37	Sensitivity to error in Floorslab density: Back wall heat flux prediction error responses
Figure 5.38	Sensitivity to error in Floorslab density: Relative contributions of driving forces to prediction errors
Figure 6.1	Classification of test facilities by level of realism
Figure 6.2	Section through an EMC test room
Figure 6.3	Adjustment of predicted floor surface temperature and heat flux
Figure 6.4	Predicted and measured zone temperature: Double glazed room with heat source at rear
Figure 6.5	Zone temperature prediction error: Double glazed room with heat source at rear
Figure 6.6	Predicted and measured floor temperature: Double glazed room with heat source at rear
Figure 6.7	Floor temperature prediction error: Double glazed room with heat source at rear
Figure 6.8	Predicted and measured floor heat flux: Double glazed room with heat source at rear
Figure 6.9	Floor heat flux prediction error: Double glazed room with heat source at rear

Figure 6.10	Predicted and measured zone temperature: Double glazed room with heat source at front
Figure 6.11	Zone temperature prediction error: Double glazed room with heat source at front
Figure 6.12	Predicted and measured floor temperature: Double glazed room with heat source at front
Figure 6.13	Floor temperature prediction error: Double glazed room with heat source at front
Figure 6.14	Predicted and measured floor heat flux: Double glazed room with heat source at front
Figure 6.15	Floor heat flux prediction error: Double glazed room with heat source at front
Figure 6.16	Relative contributions to zone temperature error: Double glazed test rooms
Figure 6.17	Zone temperature prediction error responses: Double glazed room with heat source at rear
Figure 6.18	Zone temperature prediction error responses: Double glazed room with heat source at front
Figure 6.19	Relative contributions to floor temperature error: Double glazed test rooms
Figure 6.20	Floor temperature prediction error responses: Double glazed room with heat source at rear
Figure 6.21	Floor temperature prediction error responses: Double glazed room with heat source at front
Figure 6.22	Relative contributions to floor heat flux error: Double glazed test rooms
Figure 6.23	Floor heat flux prediction error responses: Double glazed room with heat source at rear
Figure 6.24	Floor heat flux prediction error responses: Double glazed room with heat source at front
Figure 6.25	Simulated and measured responses to heater power: Zone temperature response
Figure 6.26	Simulated and measured responses to heater power: Floor temperature response
Figure 6.27	Simulated and measured responses to heater power: Floor heat flux response

CHAPTER 1

INTRODUCTION

Computer models are increasingly being used to predict the thermal performance of buildings. For a new building they may be used at the design stage to determine aspects of the design, or to ensure compliance with regulations. When existing buildings are to be refurbished they may be used to compare rival upgrade options. On a national level they can give guidance as to the energy saving potential of chosen alternative strategies.

In each of these applications it is clearly important that the models used produce reliable predictions, and that they have credibility. If not it will be impossible to use their predictions at the design stage without facing the risk of litigation. If the predictions are used at all it will only be after large safety factors have been incorporated, reducing the value of the model as a predictive tool. In refurbishment a defective calculation method may lead to the choice of an inappropriate upgrade strategy. The potential for disaster escalates when decisions which have the potential to affect large numbers of people are based on simulation modelling, as in national programmes.

In view of these considerations it is clear that simulation models need thorough testing before they are used. A range of exercises can be performed to build confidence in the programs. Broadly speaking the testing process can be divided into two categories, verification and validation. The relationship of these two activities to the model development process is discussed in Chapter 2 of this thesis. Verification addresses the problem of determining that an implementation of a model satisfactorily carries out the calculations that its author intended. In the case of large computer programmes this is not a simple task: indeed it is now acknowledged that for large programs it may be impossible ever to achieve total confidence [Littlewood and Strigini, 1992]. Many building thermal simulation codes are now large enough to fall into this category. Validation deals with the altogether more difficult problem of determining whether a program adequately represents the physical reality which it seeks to model. In order to validate it is necessary to have detailed knowledge of the physical reality which the model seeks to represent, and to know the purpose for which the model is intended: performance which is satisfactory in one context may not necessarily be adequate in another.

A wide range of techniques are now available to the model tester. Again these are classified and discussed in Chapter 2. One of the most important is empirical validation, in which the predictions of the model are compared with measurements of the corresponding reality. The appeal of empirical validation as a means of increasing the credibility of a building simulation model is immediately apparent. To represent a real building any model inevitably makes a great many simplifications. Some of these simplifications may be made in the interests of, for example, simplifying data entry or increasing computational efficiency. The impact of these approximations can be assessed by comparison with an alternative model which does not make them, and is therefore believed to be a better representation of reality. Other simplifications, however, are made of necessity rather than convenience. The underlying physical processes may be so complex that they cannot be modelled thoroughly, or that the uncertainties associated with attempting to model them are so large as to render the results useless. Empirical validation represents the only way in which the impact of this second type of simplification can be assessed.

To carry out satisfactory comparisons between data measured in the real world and the predictions of a model the empirical validator requires two types of tool: criteria for determining whether a model is performing acceptably in a given comparison, and diagnostic tools to determine the reasons for poor performance. From the review of past validation work in Chapter 2 it is concluded that there is a wealth of criteria for establishing validity, but that there are relatively few well tested diagnostic tools.

Chapter 3 of this thesis outlines the development of a new diagnostic tool for use in the validation of building thermal models. In Chapter 4 the implementation of the tool is described in detail, together with the software testing (or verification) process which has been applied to that implementation. In Chapter 5 the method is tested by using a proprietary building simulation model, SERI-RES, to generate a quasi-truth data set and then generating a series of other data sets by changing selected input parameters. The new method is shown to be capable of identifying those parameter changes, implying that when the truth model consists of data

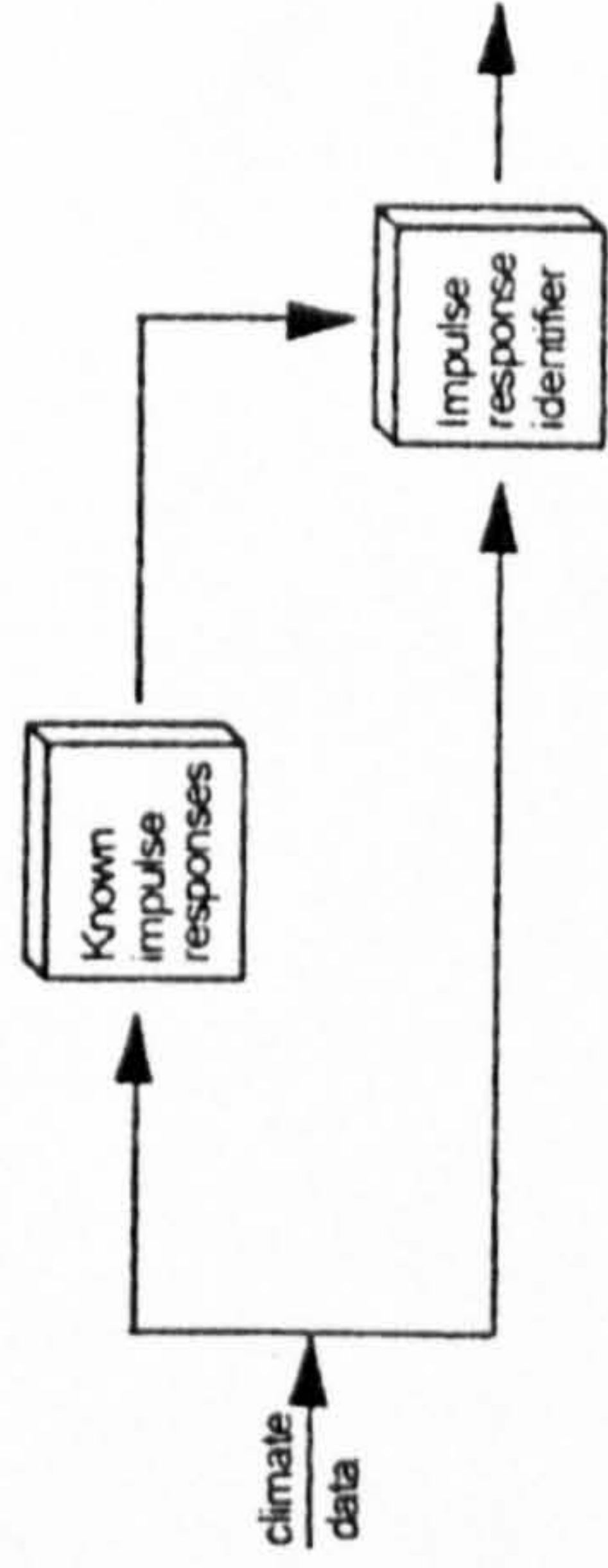
of identifying those parameter changes, implying that when the truth model consists of data measured in real buildings it will be able to pinpoint the areas which are causing any discrepancies observed between model predictions and measured reality, and thus fulfil its diagnostic potential. The principal application of the tool is identified as empirical validation, although possible applications to other model testing techniques, such as inter-model comparisons, are briefly discussed.

Preliminary validation experiments will naturally seek to eliminate as many sources of complexity and uncertainty as possible. In this way the chances of obtaining satisfactory agreement between model and data are increased, and if such agreement is not obtained there is a greater chance of finding out why. In Chapter 6 it is argued that this simplification can largely be achieved by the appropriate choice of test building. Inevitably there will be a compromise in the choice of building - too complex a building may defeat the aims of validation, but a structure which is so artificial as to be unrepresentative will mean that the findings of the project will lack credibility with real building practitioners. In the past, individual test rooms have proved to be useful vehicles for empirical validation. Uncertainties due to occupancy, complex interactions with other building zones, infiltration and ground floor heat loss can be eliminated. At the same time such rooms can be exposed to real climate, can be of a representative size and can retain realistic construction details.

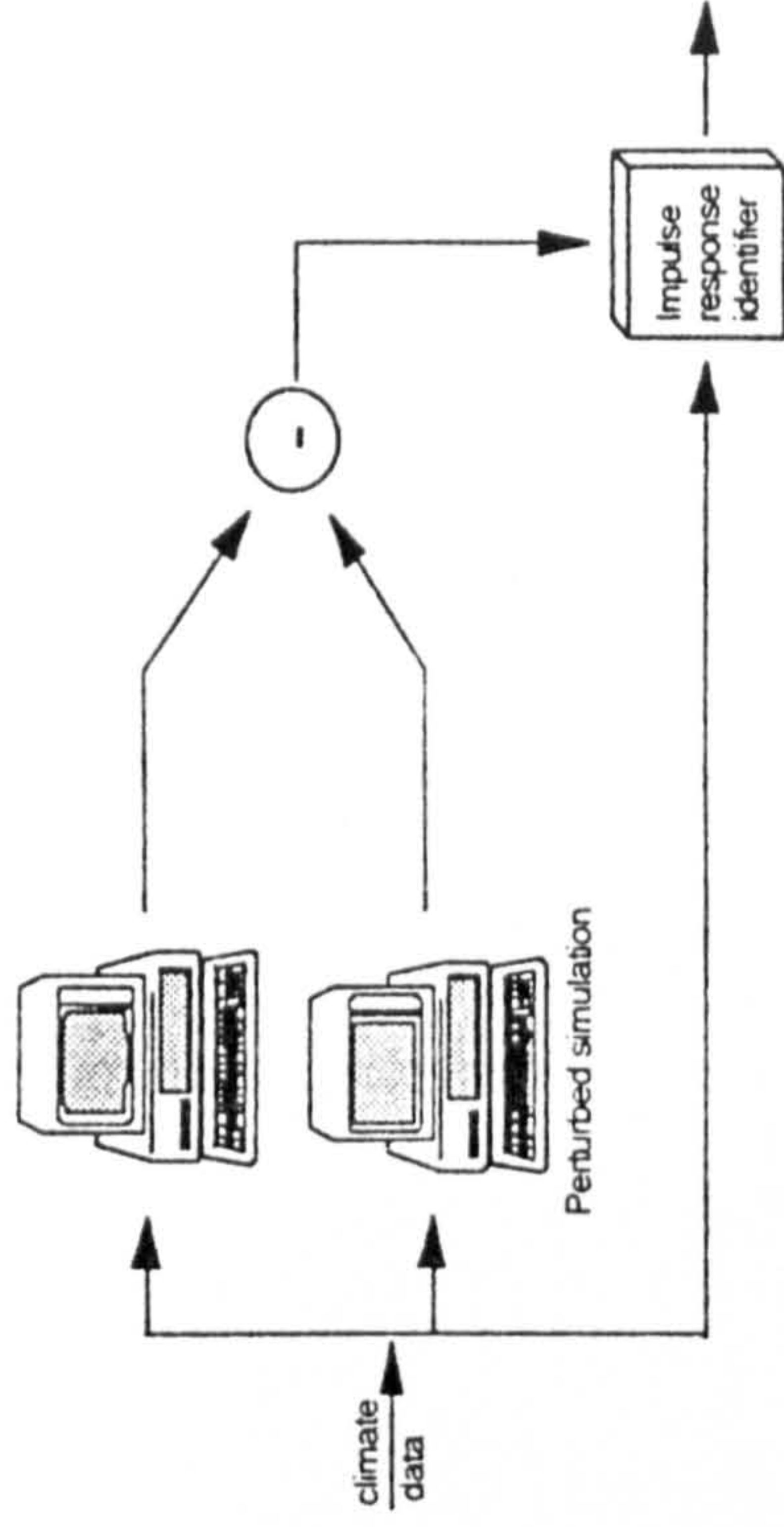
In Chapter 6 the new diagnostic tool is applied to the problem of determining the causes of discrepancies between the predictions of SERI-RES and measurements made in a pair of outdoor test rooms. The method proves capable of identifying the cause of discrepancies between simulations and measurements, which is the inadequate treatment of the interaction between the heat source and test rooms. Finally, in Section 6.8 of Chapter 6 an alternative way of applying the method is presented, which allows a generalised comparison between simulation model predictions and measurements to be made.

Figure 1.1 shows the process of proving the new technique graphically, and demonstrates clearly the progression through the various stages of testing to its ultimate applications.

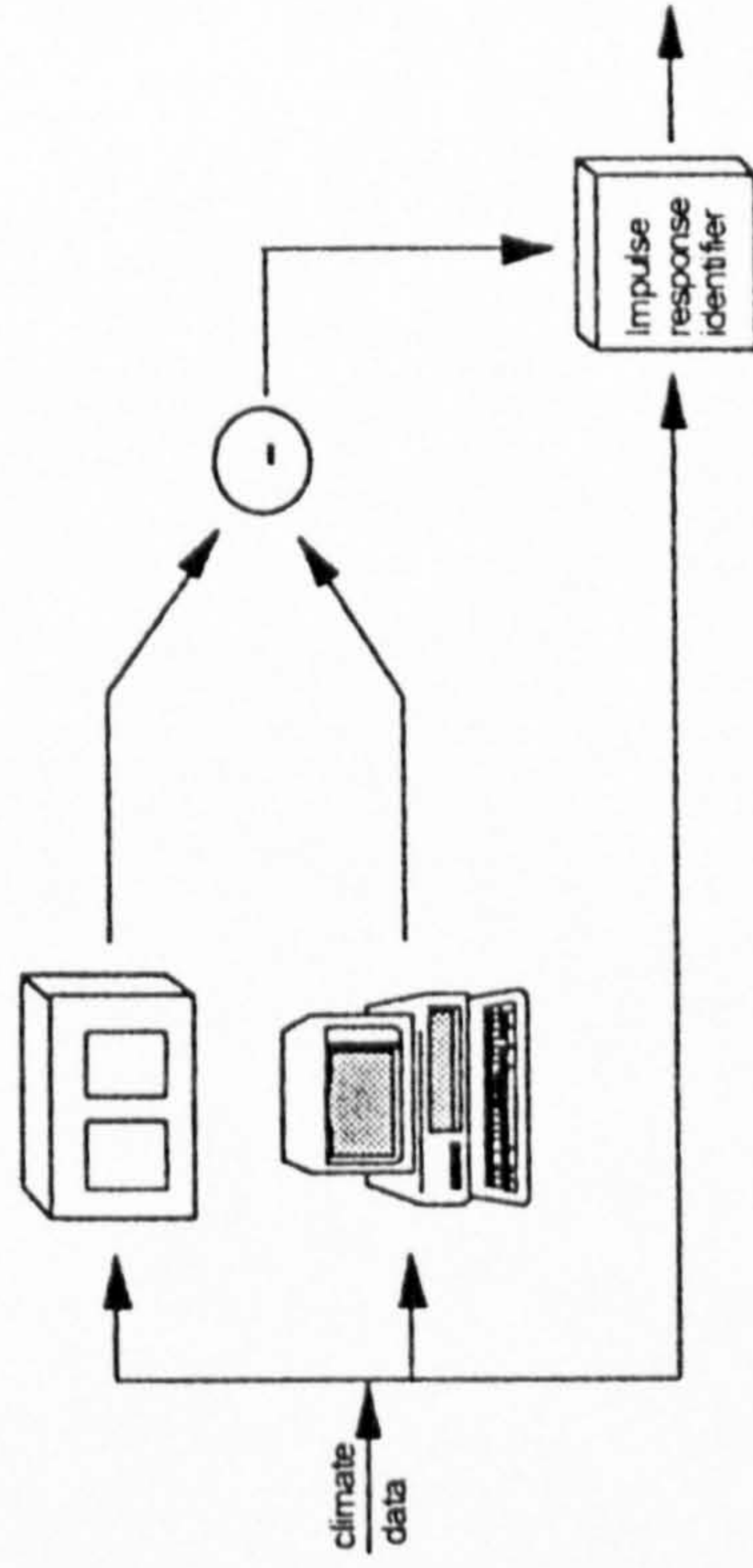
FIGURES



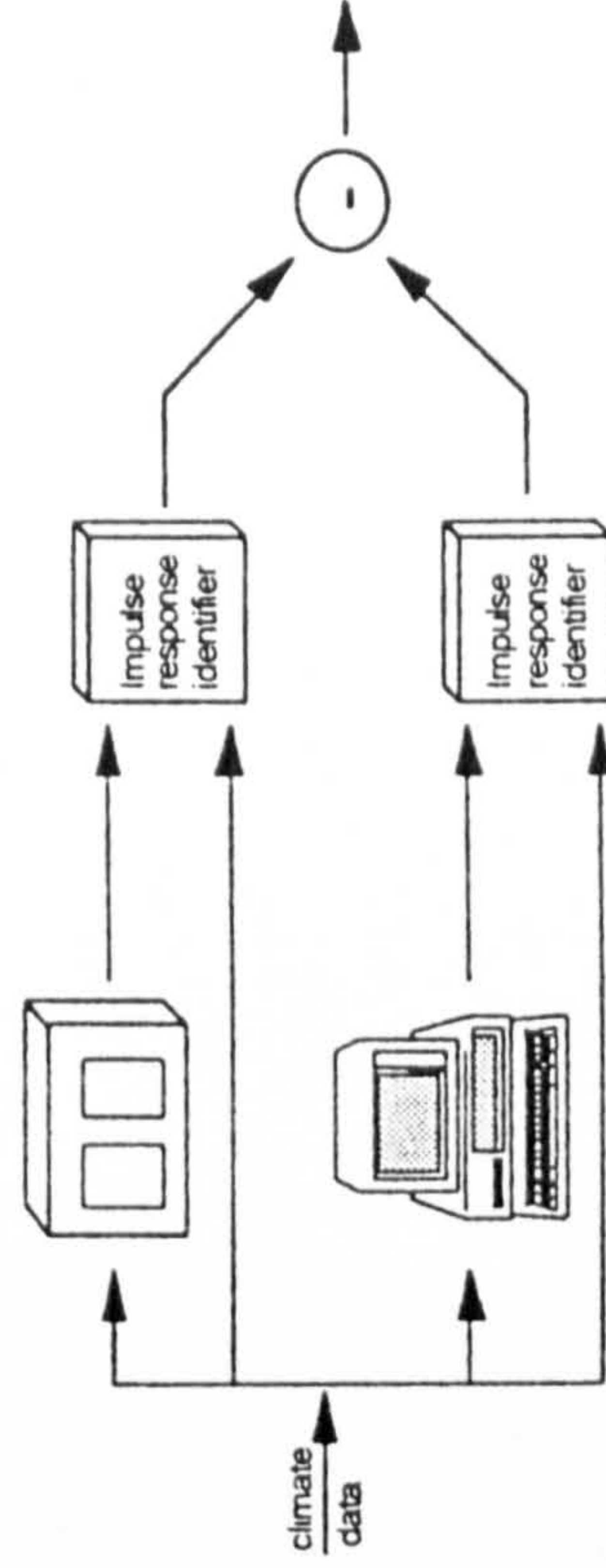
Chapter 4: Test impulse response identifier



Chapter 5: Test whether impulse identification can identify causes of discrepancies between simulations



Chapter 6: Use impulse response identification to identify the causes of discrepancies between simulations and measured reality



Chapter 6.8: Use impulse response identification as a way of comparing the response of a simulation model with measured reality

Figure 1.1: Testing and application of the new method

CHAPTER 2

A REVIEW OF MODEL VALIDATION

This chapter begins by reviewing the nature and methods of model validation. The discussion is not limited to the thermal simulation of buildings: examples are taken from fields as diverse as management science and ecology. Even so a series of common strands about the nature of validation emerge. They allow a framework to be developed within which validation can be viewed, and the tools available to the validator classified. The historical evolution of a validation methodology for building thermal simulation models is then charted. The tools currently available to the validator are discussed in two categories: criteria for establishing validity and tools for diagnosing the cause of shortcomings. Although there is a wealth of tools for establishing whether a model is valid, there are relatively few diagnostic techniques for determining the reasons for discrepancies observed when the predictions of simulation models are compared with data measured in the physical world. The techniques currently available are reviewed, and it is noted that few are ideally suited to building applications. In particular they generally discard significant amounts of information as part of the comparison process, and seldom provide diagnostic information in a form which can be readily interpreted.

2.1 What is model validation ?

It can be argued that data analysis is distinguished from the straightforward production of descriptive statistics by the fact that a model underpins all data analysis. That model may be no more complex than a belief that the ratio between two measured quantities should be constant, or it may be a detailed functional relationship relating an observed variable to a large number of input variables. In the context of such data analysis validation consists of establishing whether the chosen analysis model adequately represents data which has been collected. In this way the adequacy of the data analysis, and its associated model, can be confirmed.

The validation of models which are to be used for prediction, as opposed to describing existing data, is a more profound undertaking. The validator then seeks to address not only the issue of whether the model represents existing data well, but also the extent to which it is likely to represent future situations. Throughout the remainder of this discussion it will be assumed that

the validation process is to be applied to models which are to be used in such a predictive capacity, and these models will be termed simulation models.

Model validation is the process by which a model of a real world system is declared 'valid'. The Oxford English Dictionary [Onion, 1973] defines the word valid, when applied to an argument or assertion, as:

'... well founded and applicable, sound and to the point, or against which no objection can fairly be brought.'

Ljung [1987] proposes that the aims of the validation process as applied to models can be summarised by three questions:

- does the model agree sufficiently well with the observed data ?
- is the model good enough for my purpose ?
- does the model describe the 'true system' ?

Anscomb [1967] has taken a more radical approach to the use of the word valid:

'... the word valid would be better dropped from the statistical vocabulary. The only real validation of a statistical analysis, or of any scientific enquiry, is confirmation by independent observations.'

The idea that it is only by comparison with measurements from the real world that any model can ultimately be validated is one which will reappear throughout this review of model validation. Fortunately there are many tools which allow validation tests to be conducted without full recourse to comparison with measured data, although in many cases this is only because those tools themselves encapsulate the necessary information about reality, and by so doing allow the three questions raised by Ljung to be addressed.

Law and Kelton [1982] have presented six general perspectives on the problem of validation. They stress that this is intended not as a guide to how a model should be validated, but rather as a list of considerations about the nature of validation which the validator should keep in mind:

1. Experimentation with a model is a substitute for experimentation with the real system that the model represents. Thus validation needs to determine that experiments with the model will yield the same conclusions as experiments with the real system would, if such experiments were feasible.
2. A simulation model is only ever an approximation to the real system and will therefore never agree perfectly with the behaviour of that system. Thus validity is not an absolute quantity but rather a measure of the degree of agreement between the two.
3. A model should always be developed for a specific purpose, and a model which is good (or valid) for one task may not be valid for an alternative investigation. Thus validity must be placed in the context of application. This is further discussed in Section 2.1.2.
4. A simulation model should be validated with respect to the given purpose for which it is ultimately to be used.
5. Validation should be carried out throughout the development of the model, rather than as a product testing exercise when the model is complete.
6. The use of formal statistical tools is only a part of the validation process. Much validation will be subjective.

2.1.1 The distinction between verification and validation

Mihram [1971] introduces a distinction which will play a central part in the development of a framework in which to view model validation. He proposes that the validation process should be separated into two quite distinct activities, verification and validation.

Verification is the process of determining whether the implementation of a model (often a computer program) corresponds to the intended operation of the model, summarised by the question: Are there errors in the program? Verification is the testing of model implementation to determine whether it fulfils its specification, or that it correctly represents the collection of chosen algorithms.

The term validation, on the other hand, is applied to the process of determining that the model

is a meaningful and accurate representation of the real system, or alternatively the extent to which the model is fit for the purpose for which it is actually intended. Of the two tasks, this is more complex: it seeks to determine whether the model provides a useful representation of the observed real world system.

This is in some senses an unfortunate choice of terminology: the process of declaring a model valid, which might itself normally be referred to as validation, has been subdivided into two parts one of which is termed validation. One way around this problem is to use the term 'model testing' to describe the overall process of determining the validity of a model. Model verification and validation then form two quite distinct parts of the overall testing process. This convention will be adopted here.

2.1.2 The importance of context in model testing

Shannon [1975] opens his discussion of model validation with the warning that:

'A model should only be created for a specific purpose, and its adequacy or validity evaluated only in terms of that purpose.'

Thus the notion that validation is context sensitive is introduced. Shannon goes on to classify the validation process in terms of the steadily increasing cost of carrying out validation, and the steadily diminishing increase in validity associated with that cost.

Law and Kelton [1982] extend the notion of validation context, insisting that a model should always be developed with a clear definition of its ultimate purpose, and then validated in the light of this purpose. They warn that a model which is valid for one task may not necessarily be suitable for another.

Finally, Sage [1982] proposes a definition of validation which is centred on context:

'Validation is the process of checking the worth of a device or model in relation to its intended purpose and the system it represents.'

The issue of validity is clearly closely linked to the application to which a model will be put, and validity for one purpose does not necessarily imply validity for another. Thus the model validator should not be seeking to declare a model universally 'valid', but rather to demonstrate its validity for a given application or range of applications. The process of declaring a model valid can be viewed as one of progressively expanding the range of inputs over which the model is known to function acceptably. When a region is discovered in which the model does not function adequately that area can be delineated as one where the model is inappropriate, or the model can be enhanced to allow its range of validity to be expanded. The subsequent rectification of these errors will certainly enhance confidence in the model, and hence its validity. From these considerations it is clear that model testing is a progressive activity, each verification or validation exercise expanding either the range of applications for the model, or the range of inputs over which the model can be used with confidence.

Caswell [1976] reviews the validation problem in the context of ecological modelling, and in doing so cites an example which provides a demonstration of the importance of context in the model testing process. Using an example from population modelling [von Foerster, Mora and Amiot, 1960] he demonstrates the crucial difference between the validation of explanatory and predictive models. The Foerster model of world population growth has been shown to match observed population data extremely closely from 0 AD to the present: it is thus currently considered valid. However, inspection of the form of the model shows that it predicts that the population reaches infinity within a finite time. Foerster estimates this time (he refers to it as 'Doomsday') as 2026.87 AD \pm 5.50 years. Caswell explains this apparent contradiction, that a 'valid' model can produce predictions which are clearly untenable, in terms of the distinction between models for understanding (or analysis) and models for prediction. In fact the apparent paradox can also be explained using the idea of validation context developed here. The model can only be deemed valid over the range for which it has been validated. Thus the population model is only valid over the period 0 AD to the present, and cannot be considered valid, or used with confidence, outside this range.

2.1.3 Model validation and falsifiability

Naylor and Finger [1967] acknowledge that to validate a model means to prove it to be true, although their discussion has to be interpreted with care because they do not explicitly distinguish between verification and validation. In view of the difficulty of establishing rules or criteria for determining whether a model is 'true' they suggest that instead the credibility of a model is allowed to grow whilst no negative results are found, with confidence increasing as more and more cases which the model can represent are found. This is closely related to the theory of science advanced by Popper [1959], in which theories are accepted as true until they are falsified.

Vemuri [1978] takes this argument further, and asserts that the process of validating a model is in fact no different to the validation of any scientific theory. He does concede, however, that the degree of complexity associated with many models requires a different practical approach to the process of validation. Vemuri suggests that the validity of a theory or model may be tested by posing a series of questions:

- does the theory (or model) permit careful and accurate classification of facts and observations ?
- does the theory provide scope for discovery of scientific laws by creative imagination ?
- is the theory equally valid for all normally constituted minds ?
- is the theory capable of withstanding criticism ?

The final question leads Vemuri to describe the validation process in terms of the Popperian notion of falsifiability. A theory or model is only 'scientific' if it is capable of being shown to be incorrect. Until it is shown to be incorrect it cannot be considered invalid. Whether this allows a model to be considered valid again depends on the application context: if it is to be used for a critical task it is unlikely that the mere fact that a test has yet to show it to be inadequate would be considered a sufficient demonstration of validity.

Sargent [1982] again asserts that in validating a simulation model one does not prove the model is adequate under all sets of conditions. Instead, he advocates validation as a confidence

building exercise. This approach is complementary to the 'valid until falsified' approach introduced above. In this light, model validation is viewed as the process of successively failing to falsify a model. A model is considered 'valid' until physical measurements are obtained which it fails to reproduce. Until this happens, the area of application of the model discussed in the previous section is progressively expanded. When a falsifying dataset is encountered that area of application is delineated.

Implicit within this view of validation is that tests aimed at invalidating a given model must be the most severe which can be generated. This idea forms a central thread in the approach adopted to the design of validation experiments throughout this work.

2.2 Approaches to model testing

In this section the techniques available to the model validator are discussed. Techniques for testing models are, however, closely linked to model development, and some reference to the overall development process is therefore inevitable. Indeed, the whole model proving process will ideally be closely integrated with the development of the model, with model testing and improvement forming a cyclic process. With this aim in mind it is clearly useful to view model testing from the standpoint of the model developer.

One of the first discussions of how to conduct the testing of computer simulation models is given by Naylor and Finger [1967]. They began by postulating three fundamental approaches to the problem of modelling:

- rationalism, where it is assumed that everyone knows whether the assumptions underlying the model are true, and that the model thus consists simply of logical deduction from a series of unquestionable truths,
- empiricism, which requires that every assumption and outcome of the model be empirically confirmed, this being the sole source of knowledge about a system, and
- positive economics, which takes a more pragmatic view, and requires only that the model can predict the future, with no concern as to how it achieves this end.

Naylor and Finger combine these into an approach to model testing which they term multi-stage verification. Under the assumption that each approach will be adopted at some stage as a model is developed, the validation process has to address each approach. The multi-stage process contains one element corresponding to each of the approaches to model development:

- developing the model's assumptions,
- validating the model assumptions by empirically testing them, and
- testing the model's ability to predict the behaviour of the system under study.

Shannon [1975] considers a simulation model to be a theory, describing the structure of a system. He elaborates the ways in which the producers of such models can view the world, and hence the activity of validation:

- subjective vs objective methods: when building models there is a constant conflict between the need to be objective, and the need to make use of insights, intuition and opinion. Shannon suggests that this apparent conflict is resolved by recognising that the model development and testing process involves constantly switching between the two views. In particular he asserts that the model construction process is likely to be highly subjective, but that model testing should be an objective process.
- rationalists vs empiricists: as discussed above, the rationalist considers that a model consists only of deductions from a set of assumptions which are so clearly true as to be obvious. The empiricist, however, refuses to admit any premise which cannot first be shown to be true by the analysis of empirical data. When it comes to validation, Shannon asserts that as soon as a rationalist attempts to spell out the assumptions which underlie a given model their 'obviousness' becomes questionable. The empiricist, on the other hand, acknowledges that the validity of a model springs from its ability to reproduce observed results.
- absolute pragmatists: the pragmatist sidesteps the arguments of the rationalists and empiricists, and accepts any model which fulfils the required purpose, regardless of how it has been derived. The pragmatist is not concerned with the validity of the internal assumptions or workings of the model, merely with the input-output relationships which characterise the model. A model is more valid than another if it produces superior

predictions.

- utilitarians: Shannon argues that one seldom comes across a pure rationalist, empiricist or pragmatist. Most modellers take elements from all three approaches. This is termed utilitarianism, and forms the basis of Shannon's approach to validation.

The utilitarian approach to validation is again expressed in three stages:

- the first stage is to seek face validity of the internal structure of a model. Shannon argues that most complex simulation models consist of a series of linked sub-models of a large number of simple processes. The first stage of validation consists of checking that these building blocks are the most appropriate,
- the second stage consists of empirically testing these sub-models, and finally,
- the third stage of Shannon's approach to validation is to ensure that the model can predict the behaviour of the complete system. The stated aim of this stage is to convince the user that the model can do what it claims to do.

Law and Kelton [1982] expand the approach originally proposed by Naylor and Finger [1967] into a three stage approach to model testing, the stages of which echo exactly the approach developed by Shannon:

- develop a model with high face validity
- test the assumptions of the model empirically
- determine how representative the simulation output data are.

The weakness of all of these approaches lies in the third stage. Recall from Section 2.1 that Ljung [1987] summarised the aims of the validation process by three questions:

- does the model agree sufficiently well with the observed data ?
- is the model good enough for my purpose ?
- does the model describe the 'true system' ?

Ljung asserts the philosophical impossibility of ever determining the answer to the third question, although he gives no explanation as to why. One way of viewing the problem is to consider that in order to determine whether a model describes the 'true system' it is necessary

to have a full description of that system: in short a fully validated model. The process of validating a model by this route is therefore impenetrable: validation cannot be achieved until it has been achieved. Unfortunately, it is just this question which the third stages of Shannon's and Law and Kelton's approaches require the validator to answer. An alternative approach which initially seems more attractive is to first determine that the sub-models contained within a large model are in some sense 'valid', and then to determine that those modules have been correctly linked together. From the twin assertions that the modules are individually valid, and that they are correctly linked it might then be inferred that the complete model is 'valid'. The problem with this approach, however, is that the range of application over which the whole model can be considered valid is the range of operating conditions over which all the sub-models are within their respective ranges of applicability. For a model consisting of many sub-models with complex interactions between them it might clearly be very difficult to determine what the range of applicability of the whole model actually is.

2.2.1 Testing in the context of a three stage modelling process

It has been concluded that validation schemes which require that the validator establish whether the model is a true representation of a physical system are unlikely ever to be successfully implemented, because they inevitably require full knowledge of the system: in effect a previously validated model. In this section an alternative approach is presented, in which the modelling process is broken into three distinct stages, and the verification and validation of each of these stages tackled separately. Hoover and Perry [1989] divide the model development process into three stages:

- forming a conceptual model, which consists of identifying the parts of the real system which should be included in the model,
- forming a logical model, which consists of deriving the logical relationships between the elements of the model and the exogenous variables which affect the system, and
- generating a computer model which executes the logic defined in the previous stage.

They make extensive use of the twin concepts of verification and validation introduced in Section 2.1.1, describing how these separate processes should be applied to all three of the

modelling stages. Table 2.1 details the issues which arise when verifying and validating each stage of the development of a simulation model.

Model	Verification	Validation
Conceptual model		Does the model contain all relevant elements, events and relationships ? Will the model answer the questions of concern ?
Logical model	Are events represented correctly ? Are mathematical formulae and relationships correct ? Are statistical measures formulated correctly ?	Does the model contain all events included in the conceptual model ? Does the model contain all the relationships of the conceptual model ?
Computer model	Does the code contain all aspects of the logical model ? Are the statistics and formulae calculated correctly ? Does the model contain coding errors ?	Is the computer model a valid representation of the real system ? Can the computer model duplicate the performance of the real system ? Does the computer model output have credibility with system experts and decision makers ?

Table 2.1: Issues addressed in model verification and validation
(from Hoover and Perry [1989])

The benefit of this progressive process is that, although the question as to whether the computer model is a valid representation of the real system appears in the table, the validator is not required to address this in a single step. Instead, the suitability of the whole model for any given purpose is established by verifying and validating each of the three stages in the modelling process in turn.

2.2.2 Tools for model testing

Sargent [1982] provides a comprehensive list of validation tools, a total of thirteen techniques which can be used in the course of verification and validation:

1. inter-model comparisons can be made with the predictions of established models which have already undergone some validation. They may serve to build confidence in a new model. However, this relies on the prior existence of models with a suitably high degree of credibility with which to make the comparisons,
2. degenerate tests consist of removing selected parts of the model (either literally or by the appropriate choice of input parameters). They determine whether the behaviour of the model is correct in such simplified situations,
3. event validity tests compare the occurrence of events within the model with those in the real system, and thus determine that events are correctly represented,
4. extreme condition tests exercise the input values of the model to their extreme values, and confirm that the model behaves appropriately for these limiting points in its area of applicability,
5. establishing face validity consists of determining whether the model is 'reasonable', and also ascertaining whether selected input/output relationships are acceptable, resulting in increased credibility for the model,
6. historical data validation consists of using existing datasets to confirm the correct function of a model. This is referred to as 'empirical validation' in this work,
7. internal validity tests determine the variability between successive runs of a stochastic model. Excessive variability will lead the model user to question the results of the model,
8. multistage validation attempts to combine the three approaches of historical validation previously identified by Naylor and Finger (rationalism, empiricism and positive economics) into a multi-stage process:
 - developing the model's assumptions
 - validating the model assumptions by empirically testing them
 - comparing the input/output relationships of the model to the real system.
9. graphic displays can be used to display the model's behaviour as a simulation is carried out. Such displays may provide a useful indication as to whether the model is behaving as it should,

10. parameter variability (sensitivity analysis) determines the impact of variations in model inputs on the resulting predictions. This technique provides information which has application in at least three contexts:
- it determines the input parameters which are the most important, and which therefore need to be known most accurately before using the model,
 - the correct values of selected sensitivities may be obvious from considerations of the physical system which underlies the model. If the results of sensitivity analyses do not accord with these values errors clearly exist within the model,
 - it will be seen later that a knowledge of the sensitivity of a model to its input parameters allows confidence intervals to be placed on the predictions of the model, and that this forms the basis of one class of techniques for determining whether a model is valid.
11. predictive validation seeks to determine whether the model can predict the performance of the real system, and seeks to establish whether the predicted and observed behaviour are in some sense 'the same'.
12. traces follow the behaviour of selected parts of the model throughout a simulation, and allow the validator to determine whether their behaviour is consistent, and finally
13. Turing tests consist of tests in which a panel of experts are required to discriminate between the predictions of a simulation model and measured reality. If the experts cannot distinguish the results of the model then the model is deemed to be valid.

It is clearly tempting to try to assign to each of these tools a unique position in the process of verifying and validating the three modelling stages identified in the previous section. In fact this results in considerable overlap, with many of the techniques finding application at several stages in the model testing process.

Ljung [1987] presents a range of tools which can, in principle, be used to address the first two of his validation questions. These are:

- determining the feasibility of any physical parameters identifiable within the model

- checking the consistency of the model input/output behaviour
- using model reduction techniques to determine whether parts of the model are redundant, and can therefore be removed
- tests for statistical independence of model residuals (the discrepancies between predicted values and measured data). This forms the basis of a test for whether a given level of agreement between model and measurement implies that the model is valid, a topic examined in more detail in Section 2.3 of this chapter.

Hoover and Perry [1989] again list a series of techniques, and do make an attempt to classify them in terms of their application to each of the three stages in the modelling process presented in the previous section:

Validating the conceptual model:

- event representation of the system
- explicit identification of the elements to be in the model
- Verifying and validating the logical model:
- verifying and validating the processing of events
- verification of formulae and relationships
- verification of statistics and measures of performance

Verifying and validating the computer model:

- structured programming methods
- tracing the simulation to ensure that it evolves in the way intended
- program testing
- logical relationship tests
- comparison to analytic models
- graphical exploration of the structure of model output.

A wide variety of tools available to the model validator has been described in this section, and placed in the context of a systematic approach to verifying and validating the separate stages of the modelling process. It is clear that many of these techniques require the validator to have

full access to the theoretical framework which underlies the model, and to the details of the computer implementation which realises the model. If comprehensive testing is to be attempted on a model for which this information is not available then many of the tools described will not be usable. In this case the process of declaring the model valid will be slower and less effective than when the internal workings of the model are accessible.

2.3 Validation of building thermal simulation models

There are many thermal models of buildings, and many more buildings in which data for the validation of those models could conceivably be collected. Not surprisingly then, there have been many attempts to validate these models using data measured in real buildings. A literature search carried out in 1985 [Lomas, 1991] revealed some 600 datasets available for empirical validation. However, it was concluded that of these, only 27 were from experiments conducted in such a way as to be useful in actually identifying defects within the model being tested.

In this section only building thermal model validation projects which have contributed to the development of validation methodology are reviewed. A few individual validation experiments are also referred to, but only when they allow general conclusions to be drawn as to how validation should, or should not, be conducted.

An early study of model validation methodology for heating and cooling system simulation models was carried out in the US [Anand, Kennish, Knasel and Stolarz, 1979]. This study produced a four level methodology which centres on the division of models into two classes: detailed models and simplified models. The four levels are:

Level 1: Validation of detailed simulation programs. It is acknowledged that even detailed simulation models contain, of necessity, simplifications and approximations. Examples which are cited include lumping of parameter values, constant flow rates, spatial uniformity of the environment, lack of treatment of transients, use of constant parameters and stratification. The question addressed at this level is: What effect do unmodelled physical phenomena such as these have on the accuracy

of simulations ?

- Level 2:** Inaccuracies introduced in simplified analysis techniques. At this level the impact of the further approximations introduced when a simplified modelling tool is developed has to be assessed. This can be done by comparisons with the more detailed models considered at level 1.
- Level 3:** Variation in results due to uncertainties in input parameter values. At this level the errors introduced by the model user, who in using the model supplies parameters which are necessarily uncertain.
- Level 4:** Finally, level 4 seeks verification of the level 3 conclusions by comparisons between the predictions of the simplified models with field data.

A case study is presented which addresses levels 2 and 3 in some detail using Monte Carlo Analysis, a technique described in more detail in Section 2.3.4. However, little consideration is given to the fundamental problem of validation described at levels 1 and 4. The principal problem with this methodology seems to lie in the rather artificial distinction between detailed and simplified models. All models of necessity make some approximations (indeed, some of the simplifications included in many detailed models have been listed) and thus the distinction between detailed and simplified models is one of degree only. In particular, as more detailed models are developed, models which are currently classed as detailed may come to be considered simplified.

A second study of thermal simulation model testing at the American Solar Energy Research Institute (SERI) [Judkoff and Wortman, 1983], [Judkoff, 1988] lead to the development of a three stage validation methodology. That methodology was summarised by the techniques which were to be used at each stage:

- inter-model comparisons. A series of comparisons between SUNCAT, DEROB, DOE and BLAST was made. These served to demonstrate that the version of DEROB under test contained errors so significant as to render it useless in any real application,
- analytical testing. A series of analytical tests was derived which tested the response of massive elements, the treatment of ventilation and the transmission of solar radiation

through glazing,

- empirical validation. The issues arising in the collection of datasets for empirical validation were discussed at length, and a number of datasets collected and documented. Some of these experiments were carried out in simple test cells. These experiments were designed in such a way that 'uncharacterisable' mechanisms were eliminated. Specifically:
 - ground coupling was eliminated from the experiment by supporting the buildings clear of the ground,
 - infiltration was eliminated by tightly sealing the buildings,
 - air gaps were eliminated from wall sections, so that all heat loss was through solid materials,
 - advection and stratification within the cells were eliminated by the installation of fans, and
 - the exterior surfaces of the cells were painted white, to minimise the absorption of shortwave radiation.

These moves can be regarded as good experimental design. Eliminating as many mechanisms as possible from the trial serves to make the exercise more sensitive to those mechanisms which remain. However, the processes listed above were actually eliminated on the grounds that they could not be represented in the models under test. This can be seen as an attempt to modify the reality normally encountered in buildings to make it better agree with the abstractions inherent in the model under test. In this case, the measures serve to weaken considerably the power of any empirical validation based on these datasets. The distinction is a subtle one, and is discussed further in Chapter 6 when the choice of outdoor test rooms as buildings in which to conduct empirical validation experiments is described.

The first International Energy Agency (IEA) task addressed the problem of simulation model validation [Irving, 1982]. The study consisted of two parts, an inter-model comparison between the 23 models taking part and an empirical validation experiment. It is perhaps a sign of the optimism of that time that for the latter task the building chosen was an occupied commercial

building, the Avonbank Office block in Bristol. It was acknowledged that there were large uncertainties in at least three aspects of the operation of this building:

- the degree of convective coupling between the zones on each floor was largely unknown,
- the infiltration rate in the building was unknown, but was thought to be the dominant term in the building winter energy balance, and
- the rate of fresh air supply to each zone was also unknown.

It was noted that the considerable divergence between model predictions and measured reality could easily be accounted for by these sources of uncertainty. It was suggested that the discrepancies observed could be due to be the way in which thermal storage was treated. This was 'confirmed' by the observation that models which tended to over-predict winter heating loads also tended to under-predict summer cooling energy. It was also remarked that 'a full appreciation of the various effects causing differences is difficult to achieve because of the multi-zonal nature of the problem'.

In 1978 the UK Science and Engineering Research Council (SERC) initiated a major collaborative project with the Building Research Establishment (BRE) to study model validation [Bloomfield, 1985], [Bowman, Lomas and Bloomfield, 1978]. The project, which was part of a much wider initiative in the field of building science [Day, 1982] extensively reviewed previous validation efforts, and developed a more detailed, six stage, validation methodology, which extended the process presented by SERI:

- review of theory
- analytical verification
- inter-model comparisons
- sensitivity analysis
- empirical validation
- influence of the model user

As well as developing the basic validation methodology, the SERC/BRE project produced a wealth of work addressing the issues surrounding the validation of building thermal simulation models:

- Allen and Whittle [1986] produced detailed descriptions of the physical processes underlying external convective heat transfer and the transmission of solar radiation into enclosures. They also used an analytical approach to establish the impact the simplified representations of these processes which are used in many simulation models.
- Bland and Bloomfield [1986] produced a series of analytical tests which could be used to determine that a model correctly calculated heat flows into and within the thermally massive elements of buildings.
- Lomas and Eppel [1992] carried out sensitivity studies on a thermal simulation model and compared the results of three methods. This work is described in more detail in Section 2.4.2.
- Irving [1988] reviewed the range of statistical tests which could potentially be used in establishing the validity of time series simulators such as building thermal simulation models.
- Lomas [1991] outlined the way in which the empirical validation part of this methodology should be approached:
 - a base case prediction is obtained without reference to the measured performance with which it is to be compared. If these 'blind' predictions are within the uncertainty bands associated with the measured data then the model is deemed valid for the particular situation which has been explored. If not,
 - the total uncertainty in the predictions due to the combination of the uncertainties in all the information supplied to the model as inputs is determined. If the base case simulations differ from the measured results by less than the total uncertainty in predictions and measurements the model is deemed valid. This approach to determining validity will be discussed in more detail in Section 2.4.2. If the model fails this test,
 - the reasons for discrepancies between simulations and measurements are determined, either by reference to more detailed measured data (referred to as mechanism level data) or by using other statistical techniques.

Unfortunately, this rigorous approach to empirical validation from the SERC/BRE project was

not universally adopted. To the present day, validation experiments continue to be performed without regard for the carefully thought out conclusions of the SERC/BRE project. This often results in badly designed experiments which do not have the power to demonstrate that a model is invalid. Whilst this state of affairs might be agreeable to the model developer, it results in tests which, because they have no power to falsify, do not actually expand the region of validity of the model.

Consider as an example an attempt to validate a model of a vented Trombe wall, attached to a test cell in Switzerland [Basler and Hofmann, 1985]. The choice of a test cell was clearly a wise one, allowing the thermal performance of the element under test to be isolated from all the rival thermal processes which are occurring in a more complicated building. The data was compared with the predictions of five different models, the Swiss themselves using SERI-RES. In order to model the ventilation in the cavity of a Trombe wall SERI-RES requires that the user supply a circulation coefficient. The manual advocates the use of a value between zero and one. Upon comparison with the measured data it was found that 'values over 0.5 gave too large volumetric flow rates'. The coefficient was accordingly adjusted to 0.3 for the preparation of the final set of simulation results. After this it was concluded that 'simulation of room air temperature as well as Trombe wall surface temperatures were in good agreement with measured data'. This raises the question of what application the model is being validated for, or the validation context. The only possible application for which SERI-RES can have been considered to be validated in this case is one in which the user first has the opportunity to 'tune' selected coefficients. Such applications are relatively rare, although they do exist in areas such as energy targeting, where the future performance of a building is predicted in full knowledge of its past performance. What the model has not been shown capable of is making blind predictions such as would be required at the building design stage.

In a subsequent empirical validation experiment based in a sunspace [Baleynaud, Petit and Trombe, 1991] an interesting alternative criterion for validity was introduced. Surface temperatures were measured at a number of points over the back wall and compared with the

wall temperature predicted by a model which assumed uniform surface temperatures. When it was discovered that the predicted temperatures lay within the range of the measurements, the model was deemed valid. A moment's thought indicates the error in this conclusion. The model purports to predict an average wall temperature, and therefore its predictions should be compared to the corresponding measured average temperature (which would normally be calculated as the mean of the measured values). Further consideration reveals a major logical flaw in the methodology: the more violently the measured temperature data departs from the assumption of temperature uniformity across the surface the larger the range of temperatures which are deemed to be acceptable predictions, and thus the more likely the model is to be deemed valid.

The International Energy Agency (IEA) Task 8 began in 1982 and lasted for five years [Bloomfield et al, 1988]. The task compared the predictions of a range of simulation models for a series of imaginary test buildings. Some of the test cases were complex, to establish the discrepancies between the models in realistic applications. Other cases were very simple, and attempted to isolate the reasons for those discrepancies. More relevant to the present discussion, however, was the use of empirical data from real test rooms within this task. Of the three datasets available the UK team concluded that there were too many ambiguities in one of the experimental descriptions: in one case it was not known when the door between the two zones of the test cell had been opened and closed. Measurements from the remaining two cases were compared with the predictions of the 'most promising' models. Although the project had set out to establish the performance of models in passive solar design situations, the combination of test cell design and climate in the data which was used did not exercise the ability of the models to represent buildings with direct solar gains. It was remarked that the test revealed little about the performance of the models in a high solar fraction building or, in the terms of this discussion, that the context of the validation exercise was not appropriate to the task in hand. Furthermore, only a simple graphical comparison was made between simulations and measured data, leading only to the conclusion that agreement was 'quite encouraging'.

A major model testing exercise was undertaken at Leicester Polytechnic for the UK Energy Technology Support Unit (ETSU) [Lomas, 1992]. Termed the 'Applicability Study' the exercise comprised a large inter-model comparison study between the three models ESP, HTB2 and SERI-RES. The study had no empirical validation content, and is mentioned in passing here only because it served to unearth a great many idiosyncrasies in the programs, knowledge of which has proved invaluable to subsequent empirical validators. It also served to highlight the need for continued empirical validation efforts by demonstrating that on many tasks the three models gave significantly different predictions. Although the lack of an empirical truth model made it impossible to say which (if indeed any) of the models was correct it was certainly clear that they could not all be producing correct results !

The present author carried out a series of test room trials to generate data for comparison with the model SERI-RES, again in a project for ETSU [Martin, 1991]. It was during the course of this project that the early parts of the work described in this thesis were carried out. The work developed the use of pseudo-random heater schedules to minimise unwanted correlations between heater operation and climate. In a subsequent project [Martin, 1992] the impulse responses derived from the test room studies were incorporated into the code of SERI-RES to enable an assessment of their impact on more representative prediction tasks to be made. The experimental data analysed in Chapter 6 of this thesis was gathered during these projects.

In parallel with these activities a project funded by British Gas was collecting model validation data from a heavyweight test cell [Hitchin, Delaforce and Martin, 1993]. The validation approach was unusual in that, following [Martin, 1992], it consisted of comparing only the impulse responses of the quantities predicted by the models to each of the model inputs. It will be shown in Chapter 3 that the impulse response completely characterises a linear time-invariant system. Thus if such a model can satisfactorily predict the necessary impulse responses it should, in principle, be able to make a prediction for any situation.

The CEC PASSYS project consisted of a series of identical outdoor test rooms, spread across

the whole of Europe. The project was originally conceived as a component testing exercise [PASSYS, 1988], in which passive solar building components could be tested in the same way under the climate of different European countries. The project adopted the model ESP as its simulation tool. The first phase of the project was largely spent establishing the facilities, and little empirical data was produced. Inevitably, in a project which was to gather data from outdoor test rooms, the possibility of using that data for empirical model validation was realised. A subgroup was formed within the project to deal with model validation. Due to the lack of empirical data in Phase 1 of the project the activities of the validation group centred on activities such as code checking and analytical testing, and on establishing a method for carrying out empirical validation when suitable data became available. That method [PASSYS, 1989] was based on the following elements:

- a) theoretical examination of sub-models,
- b) internal consistency checks,
- c) analytical verification,
- d) inter-model comparison,
- e) sensitivity analyses, and
- f) empirical validation.

On this basis the method adopted was:

- review of theory behind different heat transfer processes in order to investigate possible alternative equations,
- examination of the theory behind ESP and a check of the source code,
- development of a test procedure selected from (c), (d), (e) and (f) above,
- designing test cell experiments for selected topics, based on preceding sensitivity studies,
- comparison of experimental results with those calculated by ESP, and
- in parallel with the above activities, preparation recommendations on modifications to ESP.

During the second phase of the project a number of datasets suitable for model validation were produced, and were used to carry out comparisons with the predictions of ESP. It was

concluded that the data was of a lower quality than expected, and that the differences between the predictions of ESP and the measured data may have been partly caused by a wrongly measured mean air temperature in the room [PASSYS, 1995].

In 1992 the UK Building Research Establishment issued an Information paper on model validation [Bloomfield, Lomas and Martin, 1992]. This reiterated the main conclusions of the earlier SERC/BRE project, and established a mapping between the types of problem which could exist in simulation programmes and the techniques which could be used by the validator to detect them. Table 2.2 summarises the relationship.

Validation techniques	Potential sources of error in simulation models			
	Coding errors	Errors in solution technique	Inappropriate algorithms	Over-simplification
Code checking and review	•			
Analytical tests	•	•		
Inter-model comparisons		•	•	
Empirical validation			•	•

Table 2.2: Relationship between simulation model faults and validation techniques
(from Bloomfield, Lomas and Martin [1992])

This mapping served two purposes:

- it demonstrated to practitioners that there was a clear relationship between model testing activities and goals, and
- it emphasised the progressive nature of the testing process.

In the same year IEA Task 21 provided the framework for the largest ever blind empirical validation exercise [Bloomfield, Eppel, Lomas and Martin 1993]. Data from the Energy

Monitoring Company test room facility was used to test 17 different models. Several of these models were operated by more than one participant, yielding a total of 24 model/user combinations. In keeping with the SERC/BRE methodology the simulations were carried out blind. The simulation modellers were supplied with weather data, a comprehensive description of the construction of the test rooms and details of how the rooms had been operated. A telephone hot line was established to give modellers the chance to clarify any aspect of this information. To ensure fairness, any information given out on the hot line was subsequently made available to all other participants via a regular newsletter. In spite of the amount of information fed to modellers and the emphasis placed on the importance of quality control at the modelling stage, the initial blind simulation results showed a wide divergence (up to a factor of four on predicted energy consumptions). A large part of this range was traced to errors which might be classed as 'minor', for example results returned in kWh rather than MJ, although it has to be said that in a simulation exercise where there was no recourse to the results of others these errors might have remained undetected. In response to this problem, feedback was sent to each participant, suggesting areas in which they might like to refine their simulations. In order to maintain the blind nature of the exercise the researcher inspecting the results was at this stage in ignorance of the measured data. This first feedback stage served to reduce the range of results to around 1.6:1.

Due to restrictions of time, the degree of divergence between measured and simulated results that might be expected due solely to experimental uncertainties was estimated using one model only, SERI-RES. The formal methodology strictly requires that this error band be estimated with each model in turn. This approach was justified on the grounds that SERI-RES was one of the more simplified models taking part in the exercise, often using single parameters to represent complex physical processes. These parameters have uncertainties due to both uncertainty in the knowledge of the physical world, and to the approximate nature of the parameterisation process. It was asserted that the uncertainties obtained from models which contained a more explicit representation of such processes would therefore be smaller, and that error bars generated using SERI-RES represented the 'worst case'. The actual error band was

approximately $\pm 7\%$, meaning that if a given prediction fell more than 7% away from the measured data the model producing that prediction could be deemed invalid. It was immediately clear that the overall spread of the simulation results was much larger than the uncertainty band. Wherever the measured data had fallen no more than 8 of the 24 models tested could have been deemed valid. Interestingly then, the actual value of the measured data was irrelevant in assessing the overall state of the art in simulation modelling, only the estimated uncertainty and spread between results were important. Where the measured values did become important, of course, is in assessing *which* of the models tested were the best performers. In fact only 2 of the 24 models fell within the uncertainty range for all the cases examined.

In a document produced as part of IEA Task 21 Bland [1992] proposed a series of analytic tests of thermal simulation models. In these tests the models are used to make predictions for cases so simple that the 'correct' solution can be derived analytically. Appealing to the approach to the model testing process described in Section 2.1 Bland suggested that the division of model testing of thermal simulation models should be approached in two stages, debugging (or verification) and tests for assessing the accuracy of algorithms, or how well they represent the behaviour of physical reality (validation).

In the case of analytical tests Bland distinguishes between the two types of test succinctly. He argues that the first (verification) will always consist of comparisons between the computer model under test and an alternative implementation of the *same algorithms* which is acknowledged to be without error. This test establishes that the model under test correctly implements the chosen algorithms, and will not require any empirical data. Agreement will be perfect if the implementation is correct - any discrepancy indicates an error, and hence failure of the model under test. In practice serious problem arises in conducting implementation tests in this way. Every algorithm within a program that is being tested must be independently, and correctly, implemented by the accreditors. Thus to implement a test a whole program, the accreditors must effectively re-write the entire program, obviously ensuring that there are no

errors in their implementation of the chosen algorithms. Indeed if this were possible the program authors could have written the original program in this way, rendering implementation testing (verification) redundant ! In practice, it is well known [Littlewood and Strigini, 1992] that such error-free implementations are, if not impossible, extremely unlikely. One way of resolving this problem is to test the software at only a small number of points, chosen so that the validator can conveniently calculate the 'correct' solution, for example, transmission algorithms might be tested at normal incidence only. Points at which a result can be readily calculated are likely to be points which do not exercise the complete simulation code to any great extent. Whilst such tests might identify some problems in simulation codes there will be many more errors which will remain undetected, and the tests are therefore very weak. Until this problem can be overcome, implementation testing will remain a technique for carrying out limited tests on limited parts of a simulation code.

The second type of test (validation), by contrast, assesses how adequately the chosen algorithms represent the physical reality found in buildings. This class of test will inevitably require measurements of the performance of real buildings, and there will inevitably be discrepancies between the predictions of a given algorithm and measured reality - it is the magnitudes of these discrepancies which indicate the quality (or degree of validity) of the algorithm. The only exception to this requirement for empirical data in accuracy testing is likely to be when simplified algorithms are being tested against the predictions of *significantly superior algorithms*, which are believed to be essentially accurate, that is in the case where good, but normally computationally prohibitive, solutions already exist and can be used as quasi-truth models. One example, cited by Bland, is the three-dimensional analysis of heat flow, which could be used to judge the appropriateness of the one-dimensional algorithms employed in many thermal simulation codes, although even three-dimensional codes contain some approximations and therefore do not constitute a perfect truth model. Such cases (which are in fact remarkably rare in building physics) can be regarded as situations in which the necessary empirical data has already been collected and summarised in a form which is comprehensive, but computationally prohibitive in normal use.

The practical impossibility of comprehensive implementation testing and the inability of this approach to give a measure of overall program accuracy, leads inevitably back to exercises such as 'whole model empirical validation' and 'intermodel comparison' as the only methods for testing simulation tools *in their complete form*. In effect, the issue of whole program verification has to be addressed obliquely by carrying out validation exercises on the complete model. The principal weakness of this approach is that approximations in the formulation of the physical model inevitably lead to discrepancies, and thus verification can never be established without doubt. Instead only an estimate of verification is available, based on the quality of the validation results.

2.4 Tools for establishing model validity

In this section tools are discussed which address the first of the Ljung criteria detailed in Section 2.2: Does the model agree sufficiently well with observed data ? In general a validation dataset will consist of a collection of input data with which to drive the model, and a series of output data which the model should, if it is valid, recreate. Because the model contains approximations and the measurements of physical reality inevitably contain uncertainties this recreation is unlikely to be exact. The task addressed here is that of determining whether the degree of agreement obtained is adequate, implying model validity, or whether the magnitude of the discrepancies implies that the model is not suitable for its purpose. Approaches to this problem are reviewed under two headings here: statistical comparison tools and output uncertainty methods.

2.4.1 Statistical tests for model validity

Law and Kelton [1982] classify the statistical tools which are available for the comparison of model outputs with measured results into three categories:

- an inspection approach, where statistics are calculated summarising the output of the model and the real data, and the values of these statistics are compared, normally without the use of a formal procedure,

- a confidence interval approach in which statistics from many model/data comparisons are used to generate confidence intervals with which to test the hypothesis that the output of the model is the same as the measured data, and
- time series approaches in which the two datasets are analysed as time series. Standard statistical tests are then used to determine whether the coefficients describing the two series differ significantly.

All three of these classes of analysis consist of generating statistics which summarise the simulated and measured data, and then assessing those statistics to determine model validity. The statistical tools described of necessity destroy information, indeed this is the purpose of a statistic: it reduces the amount of information in a dataset to a point where meaningful comparison or analysis can be carried out. So rather than trying to compare two series of data, statistics such as the mean value of each sequence are first computed, and these are compared instead. However, in computing those means information has been destroyed, and even if the mean values are deemed to match the same cannot be said of the underlying sequences. Underpinning this argument is the whole issue of suitability for purpose or validation context: if a model is to be used only to predict mean values then a test of means can be used to assert validity. If, on the other hand, the model is to be used to predict, say, the maximum temperature within a building, then claims of validity based on a test of means do nothing to improve confidence in the model. It has not been validated in the context of maximum temperature prediction.

Reckhow [1988] presents a statistical overview of the validation of simulation models, couched in the Popperian framework of model validation as failure to falsify a model, described in Section 2.1.3. He lists the statistical tools which can be applied to the validation of deterministic models as the t-test, the Mann-Whitney-Wilcoxon test (a non parametric equivalent to the t-test), and regression and cross correlation methods. He also discusses their extension to multi-variate models, where the assessment must be based on multiple outputs.

The t-test is perhaps the simplest test which can be applied to a set of predictions and

corresponding measured data. This is a test of location: the criterion for validity is whether the distributions of simulated and measured data have the same mean value. Sokal and Rolf [1969] describe the use of the paired t-test as a tool for determining validity. Irving [1988] extends this, describing the use of the vector t-test on multiple time series.

Hsu and Hunter [1977] address the problem which occurs when applying simple statistical tests such as the t-test to auto-correlated sequences of data by first constructing autoregressive models of the simulated and measured data streams. They present formal statistical tests for establishing whether the sequences are equivalent in the mean, and whether their autocovariance structures are equivalent, using the coefficients of the identified models.

Jørgensen [1981] introduced the coefficient of variation of simulation errors, which he defined as the standard deviation of the residuals divided by the mean value of the measured quantity. This is clearly closely related to the t-statistic described above. Jørgensen also used timing error, the difference in the times at which the predicted and measured values of a variable reached their peak values, as a measure of model validity.

Other workers have described the use of alternative statistical tests. For example, an F-test can be applied to determine whether the distributions of simulated and measured data have the same variance. Clearly applying both t-tests and F-tests is a stronger test of validity than using the t-test alone. This raises the possibility that a model might pass a t-test and be deemed valid, but if subsequently subjected to an F-test it could fail. Once again, model validity is seen to depend on the context in which the model is to be used. If the sole requirement is that it predicts correctly in the mean then the t-test is appropriate. If, on the other hand, the variance of its predictions is also important then this must be combined with an F-test.

Considerable effort has been devoted to the use of regression coefficients to determine whether agreement between predictions and measurements is adequate [Cohen and Cyert, 1961] or [Thomann, 1980]. The principle is simple. Simulation results are plotted against the

corresponding measurements, and a straight line fitted to the data using conventional regression analysis. If agreement was perfect the result would be a line with an intercept of zero, a slope of one, and an r^2 statistic of one. In practice discrepancies between simulations and measurements will ensure that such a perfect result is not obtained. Statistical tests exist to determine whether the parameters obtained could have occurred purely by chance, or whether they represent a statistically significant alternative relationship. By considering a simple linear transformation this approach can equivalently be expressed as plotting the simulation error against the measured data. Thus the criteria being used to demonstrate validity is actually 'no detectable linear relationship exists between the simulation error and the measured value'. This is still a weak definition of validity: a non-linear relationship could exist. Regression coefficient methods have been applied to the problem of validating a building thermal simulation model by [Ishida and Udagawa, 1991]

An alternative approach, which again consists of modelling the simulated and measured data sequences, determines the power spectral density of the two datasets [Fishman and Kiviat, 1967]. Standard statistical tests are then used to determine whether the differences between the spectra are significant.

Leggett and Williams [1981] introduced the reliability statistic as a method of quantifying the performance of models against measured data. The quantity is defined in two alternative ways, geometrically and statistically, although the point is made that when agreement between simulation and measurement is close the two definitions are almost equivalent. The reliability statistic gives a measure of how close the predictions of a model can generally be expected to come to measurements. Although no formal criterion for validity is presented, the statistic provides a useful way of assessing rival models against a measured dataset.

During the second phase of the PASSYS project, described in Section 2.3, considerable effort was expended developing tools for comparing the predictions of ESP with measurements made in the test cells. As part of this activity a package of statistical tools for empirical model

validation was produced [Palomo and Téllez, 1991]. The package, which consists of a series of statistical tests on the differences between the model predictions and measured data, is intended to address the two issues central to the validation problem:

- is the level of agreement obtained between model and data sufficient for the model to be classified as valid ?

and if not:

- can the sources of the discrepancies be identified ?

Only the first of these questions is addressed here. Discussion of the second is postponed until the next section.

In selecting statistical tests to address the first question, the concept of a 'valid' model as the best choice from some overall class of models is postulated. From this definition the notion that a valid model must always produce residuals with no remaining statistical structure was introduced. Such a sequence of errors is often referred to as white noise, because its unstructured nature implies that all frequencies are present, in the same way as they are in white light. The reasoning behind this criterion is simple: if the simulation errors contained structure the model could be improved to account for the structured part of the error, and should not therefore be deemed valid in its current state. This is a common test of the quality of a fitted model in the control science field (for example Isermann [1981] or Ljung [1987]). In the PASSYS work a series of tests for the whiteness of simulation errors were presented in both the time delay and frequency domains. However, it was further incorrectly postulated that a model which produced white residuals was, by definition, 'valid'. Two very simple counter-examples can quickly demonstrate the incorrectness of this assertion:

- consider the case of a model which reproduces reality perfectly: a model which is, by any criterion, valid. However, an unfortunate error during the development of the model introduces a line of code which at each time step adds a large random number to all the predictions of the model. The residuals produced by the new model are white, but the model should not now be deemed valid,
- consider a model which contains an error which adds a proportion of the heater power input

directly to the predicted building air temperature. Now consider a validation experiment in which the building heat source is operated pseudo-randomly (a strategy which was in fact adopted at various stages of the PASSYS project). In this case the error in the model is undetected because the errors produced by the randomly operated heater will themselves have no apparent statistical structure.

The second pathology was subsequently addressed by developing a further series of tests on the correlations between model inputs and simulation error. Again, versions of the test in both the time and frequency domains were presented, the former on the cross-correlation function and the latter on the partial coherency spectra.

As part of the PASSYS component testing work an analysis technique was introduced which used an iterative process to adjust the parameters of a simple model of the performance of the test room. Although subsequent studies have raised serious concerns about the validity of this type of technique as a component characterisation tool [Dewson, Day and Irving, 1992] it is without doubt a direct example of the choice of the best model from a class of models. If the parameter fitting process is carried out correctly, the resulting fitted model is, using the definition employed in this work, valid: it is the best choice from the class of models considered. Interestingly, the proposed model validation method was applied to the predictions of a model obtained from such a parameter fitting exercise. The model failed the test, and was declared invalid. The implications of this for the model validation strategy, or for the reliability of the component test results derived in this way were not discussed.

Taken together, the statistical methods described provide a selection of tools for determining whether agreement between model predictions and measurements is adequate. Although some of the techniques may, under certain conditions, be equivalent, they generally use different criteria for defining what constitutes a significant difference between simulations and measurements. It is thus possible that in a given experiment some would indicate that a model was valid, whilst others would indicate that discrepancies between model and data were

significant. In this case it becomes doubly important to consider carefully exactly what the assumed criterion for validity actually is for each test.

2.4.2 Output uncertainty criteria for validity

The second class of tool for determining whether a model is valid relies on calculating simulation output uncertainty. Such methods address the issue of whether a model is valid by generating an estimate of the total uncertainty in the predictions of the model due to the uncertainties in all its inputs. If the difference between the model output and measurement is less than this uncertainty plus the uncertainty in the measurement then that error *could* be due to uncertainties in the experimental data, and the model may be valid. If not then the model is clearly flawed, and can be declared invalid on the basis of that test.

Miller [1974] addressed the use of sensitivity analysis in the validation of ecological models. He generated a single statistic (the D-statistic) which was a function of all the model outputs of interest. The sensitivity of this statistic to each of the model input parameters was derived and shown to be essentially constant (or, equivalently, the model and statistic combination was found to be linear in the input parameters). However, this was not used directly as a criterion for validity. Instead a series of experts were shown sets of simulation results and corresponding measurements and their opinions as to whether agreement was 'reasonable' were used to generate values of the D-statistic within which the model could be deemed valid (D_{crit}). Although this work did not explicitly establish a criterion for comparing simulation results with measured data it provided the basis for such methods, by developing the idea of a measure of quality of fit which could be interpreted in terms of simulation uncertainty.

Determining the sensitivity of a model to a large number of input parameters is, in general, a very difficult undertaking. If the model is non-linear then the number of simulations required to determine the effect of all possible combinations is given by the number of values explored for each input parameter raised to the power of the number of parameters. For the building thermal model applications considered here this would be typically 3^{30} , or 200 000 000 000 000

simulation runs ! Hearne [1987] has suggested one solution to this problem, in which parameters are grouped into combinations which are known to contribute together to the most extreme behaviour of the model output. In this way the number of parameters to be investigated is reduced, with corresponding reduction in the number of simulations needed to determine the required sensitivities.

For the case where a model is linear the sensitivity to each input parameter can be determined independently, and the number of simulations required is then reduced to the twice the number of input parameters under investigation plus a base case run with all parameters set to their nominal values. For the example above, the number of simulations required is reduced from 200 000 000 000 000 to 61. The linearity of the model allows the sensitivity to any combination of parameters to be determined by superposition of the results for individual parameters. This is commonly referred to as a Differential Sensitivity analysis (DSA).

Monte Carlo analysis (MCA) provides a well known alternative method of determining the uncertainty in model output, given the uncertainty in input parameters. A series of simulations is carried out in which the input parameters are randomly selected from their error distributions. The resulting set of simulation results is then analysed to determine the distribution of the outputs. Howell [1973] discusses the use of a Monte Carlo treatment of data uncertainties in thermal modelling problems, and compares it with the more pessimistic 'worst case' methods for calculating the uncertainty in the temperatures around a cryogenic storage tank on a high speed vehicle.

Stochastic sensitivity analysis (SSA) provides a further way of determining the required sensitivities [Irving, 1992]. A single simulation is conducted, during which each input parameter is continuously perturbed using statistically independent random sequences. Cross correlation of the resulting model output with these perturbations yields the required sensitivities. This method has the disadvantages that the computational effort required is very much greater than that for either DSA or MCA, and also that it requires modification to the

interior workings of the model, to allow parameters to be perturbed over the course of a simulation. At the same time, however, it provides much more information than either of the other techniques:

- in their native form the cross-correlations represent the way that the model responds to a perturbation in an input over time, or the dynamic sensitivity of the model,
- the cross-correlations can be integrated to give a single, overall, sensitivity to each parameter, and
- the resulting sensitivities have uncertainty bands associated with them, which give an indication of how linear the model is by demonstrating how much the sensitivity to a given parameter varies as other parameters are changed.

Lomas and Eppel [1992] carried out a comparison of the DSA, MCA and SSA methods when applied to dynamic building simulation models. It was found that the three methods gave consistent estimates of the overall uncertainty associated with predictions (the measure required to assess model validity in the way outlined here). This in turn indicated that the assumption of linearity required for DSA, although known to be incorrect for most thermal models, does not cause serious errors. Of the three techniques, only DSA and SSA can provide the sensitivity to individual inputs, and these were again found to be in reasonable agreement.

An alternative output uncertainty approach to the issue of quantifying validity is described by Butterfield [1981] and Butterfield and Thomas [1983]. In order to establish validity the method varies the input parameters to the model to force it to follow the measured data exactly. If the variation required takes the parameters outside their known limits then the model is deemed inappropriate, or invalid. At first glance this seems equivalent to the schemes described above where the uncertainties in the input parameters are used to determine the uncertainty in the output. However, by allowing the user to choose combinations of inputs which yield the required output the Butterfield approach effectively allows the relationship between the parameter uncertainties to be arbitrarily chosen. In, for example, the Monte Carlo approach, these relationships are fixed prior to carrying out the sensitivity study when the joint probability

distributions of the input parameters are specified. As a result, the Butterfield method is more likely to find a combination of input parameters which can explain a given discrepancy between simulations and measurements, and it represents a weaker criterion of validity than the other methods.

The validation work within the PASSYS project included the development of a computing environment in which the input parameters to ESP could be varied to establish the uncertainties in the model outputs as a function of input uncertainties, following the techniques developed as part of the SERC/BRE project. This was subsequently extended to allow the input parameters to be adjusted in order to produce the best agreement with a given output, and the results of such an exercise were presented in 1992. After over ten years, the PASSYS team had succeeded in implementing the technique first proposed by Butterfield in 1981.

2.4.3 Statistical validity criteria versus output uncertainty methods

From the above discussion it seems that there is a wealth of criteria for determining validity, and that they fall into two quite separate classes. The first of these, statistical methods, suffer from a general shortcoming: they make no allowance for the quality of the measured data or the input parameters fed to the model. Both the measured inputs which are applied to the model (parameters and driving sequences) and the measured results with which the model output is to be compared may be known very accurately, or their values may contain very large uncertainties. It is reasonable to believe that in the former case very close agreement between simulation and measurement should be expected, whereas in the latter case a larger discrepancy between simulation and measurement may be tolerated before the model is declared invalid.

Output uncertainty methods resolve this problem, and it thus seems clear that they should be preferred for determining whether a model is 'valid'. However the opposite problem to that identified for statistical tests may be encountered. Consider a model which contains a fault in its response to only one of many input parameters. Taken together the uncertainties in all the input parameters supplied to the model may be sufficient to generate confidence intervals on

the output which are wide enough to mask the single genuine error. In this situation a statistical test which searched for correlations between particular inputs and discrepancies might have revealed the presence of the error.

Output uncertainty methods never give a clear indication that a model is valid, only that it cannot be declared invalid on the basis of a given experiment. If the experiment is subsequently refined, reducing the uncertainty in some of the model inputs and hence in its output, may enable a previously acceptable model to be declared invalid. The approach therefore fits well into the Popperian framework of model testing presented in Section 2.1.3.

However, it is acknowledged that models of necessity contain simplifications, and as experiments consistently improve all are therefore ultimately destined to be declared invalid using output uncertainty criteria. It is at this stage that the context of the validation exercise becomes important: as better and better experiments reveal the shortcomings of a model it is the responsibility of the validator to delineate the applications in which it can still be used with confidence.

2.5 Tools for diagnosing the reasons for poor predictions

Early attempts at diagnosing the causes of poor agreement between models and reality relied largely on the calculation of cross-correlation functions. As part of the SERC/BRE collaborative validation project the cross-correlations between model predictions and measured results were evaluated [Lomas and Bowman, 1987]. These allowed a simple test to be made of whether discrepancies were due to time delay (or phase shift) effects. By implication it was then possible to infer whether it was the thermal mass or the heat loss of the building which was incorrectly modelled, the former being expected to cause phase shifting, the latter not.

A natural extension of such cross-correlation analysis is to calculate the correlations between simulation error and model inputs, in an attempt to determine which inputs are responsible for those errors. Such a simplistic approach, however, can give misleading results [Martin, 1990]. Because some of the inputs driving the model are themselves correlated (for example, solar

radiation and external temperature) an error in the processing of one of these variables will generate an error sequence which also correlates with the other, related, inputs. In their contribution to the PASSYS project described in the previous section Palomo and Téllez [1991] had overcome this problem by computing partial cross spectra between model inputs and simulation errors. However, only passing reference was made to the fact that the results of these tests could potentially be interpreted in terms of which inputs were responsible for the observed errors.

An alternative approach to identifying the reasons for discrepancies between simulation results and measurements uses spectral decomposition to explore the forms of both sets of data. Spectral techniques have been used in the validation of a building thermal simulation model by Candau and Piar [1991]. The technique generates a spectral map of each sequence of measurements and is able to isolate a small number (four in the case presented) of modes which largely determine the performance of the underlying system and corresponding model. The technique is used to reveal a problem in the modelling of a reinforced concrete floorslab in a test house.

Ljung [1987] makes brief reference to the idea of constructing a frequency response model of simulation error, in order to determine the frequency ranges in which a model does not represent measured data. This spectral approach has been further developed by Ramdani [1994]. Working in the frequency domain he uses spectral analysis to decompose simulation errors into a range of frequencies, and identifies which driving forces are contributing to the simulation error. However, the resulting interpretation of these results, which are in terms of the phase and magnitude of the contribution to error at different frequencies, is difficult.

It is clear from the discussion in this and the preceding section that although there is a wealth of criteria for establishing the validity of a model, there are relatively few methods for determining the reasons for any discrepancies observed between model predictions and measured data. It is further clear that the techniques which are available do not provide results which are

amenable to interpretation in terms of the physical processes known to exist in buildings. There is clearly a need for a technique which fulfils the following criteria:

- the technique should be particularly careful about retaining information which is known to be relevant to buildings, specifically dynamic information,
- the technique should destroy a minimum of relevant information,
- the technique should provide diagnostic information in a form which can be readily interpreted in terms of the physical processes which are understood to be present within the building.

The new technique which will be developed in the remainder of this thesis is one attempt to address this need.

CHAPTER 3

A NEW METHOD FOR IDENTIFYING SOURCES OF ERROR

The model validation techniques reviewed in the previous chapter all seek to characterise discrepancies between simulation results and reality in terms of the statistics of those errors. In this chapter an alternative approach is proposed, in which a mathematical model is fitted to the process or processes causing that error. Inspection of that 'error model' may then allow the contributors to the error to be identified, and the nature of their contributions to be characterised.

3.1 A linear time-invariant error model

In order to separate and identify the sources of discrepancies between the predictions of a simulation model and the reality it seeks to represent, a simple mathematical model of the error process is postulated. The proportion of the error in any predicted variable due to a given driving force is assumed to be related to that driving force by a linear, time invariant dynamic system. In reality, this assumption will always be violated to a greater or lesser extent. For example the dependence of a simulation error on solar radiation may vary over time as solar geometry changes. Other sources of error may violate the linearity assumption, for example errors in external surface heat loss may depend on wind velocity, solar radiation and net radiation exchange in a complex, non-linear, way. Alternatively, coding errors in the model may have introduced arbitrary non-linear or time varying errors in its predictions. In any real application it will be possible to test how adequately the assumptions of time invariance and linearity allow the observed errors to be represented, and at this stage a decision can be made as to whether a more sophisticated error model should be employed. Initially, however, we will concentrate on the use of a linear, time invariant model.

There are many ways of representing the performance of linear, dynamic, time-invariant systems. To display dynamic behaviour the model must of necessity include terms which make the current output of the system a function of its past behaviour. The various representations available approach this problem in different ways, and as a result some are more general than others. However, as might be expected, since they all seek to describe the same class of system, there is considerable overlap between alternative representations.

It is a requirement of any physical system that it should not respond to inputs which occur in the future - it must be causal. This is not the case for simulation errors: a coding error could introduce a non-causal error component. The models described here will all be presented in their conventional causal forms. If a simulation error is believed to have a non-causal component this can be identified by time shifting the output sequence in relation to the inputs.

In examining alternative representations of the chosen error model a series of criteria relevant to the particular application considered here will be considered:

- can the model readily be derived from measured records of system input (in this case values of the variables driving the simulation process) and output (in this case the discrepancies between simulation and reality) ?
- to what extent does the structure of the model have to be defined before it can be identified, or, equivalently, what assumptions have to be made about the data being modelled ?
- is the identified model amenable to interpretation ?

Since both the outputs of simulation models and measurements made of reality are at discrete points in time, all the system descriptions will be presented in their discrete time series forms. In each case, for the sake of clarity, systems with a single input will be considered, even though the final application clearly requires the extension to multiple inputs. That input will be denoted by the sequence $\{x[t]\}$, where t takes the values 0, 1, 2 and so on up to a maximum of n , implying that there are $n+1$ points in the dataset. The system output sequence will be denoted by the corresponding sequence $\{y[t]\}$.

3.1.1 State-space models

In the state-space representation the 'state' of the system being modelled at any given instant summarises the entire history of the system. The vector of state variables at time t will be denoted here by the vector $\underline{s}[t]$. In the case of physical systems whose dynamics are due to the storage of energy in lumped elements, for example electrical circuits, and lumped thermal

models the state vector can be chosen such that its elements represent the amount of energy stored in each such element [Owens, 1981]. The discrete time state space model takes the form of an equation for updating the state vector between timesteps:

$$\underline{s}[t] = A \underline{s}[t-1] + B x[t]$$

and an expression for generating the output of the system as a linear combination of the elements of the state vector and the system input:

$$y[t] = C \underline{s}[t] + D x[t]$$

where:

$\underline{s}[t]$ is the vector of the system states at time t , and

A , B , C and D are matrices containing the model parameters.

In the case of a continuous time system a state-space representation of finite order cannot represent a system which incorporates a pure delay. For the discrete time systems considered here this constraint does not apply: a pure delay can always be incorporated simply by extending the number of states and storing old inputs in them until they are required.

A range of identification methods exist which enable state-space models to be fitted to measured input and output data [Isermann, 1981]. It is often incorrectly asserted that the Kalman filter is one such technique. In fact the Kalman filter allows for recursive state estimation for the case where the underlying state-space system is known [Kalman, 1958]. It is possible to recast many problems so that the system model parameters also appear as unknowns, but this no longer represents a true Kalman filter. In spite of this, the technique has been found to give surprisingly good results in a range of circumstances [Gelb, 1974]. There are a number of more rigorous approaches to the problem of identifying state-space models, [Iserman, 1981]. These often have the advantage of being recursively formulated, implying that the computational requirements will remain modest as the number of data points to be analysed increases.

However, they all require that the user specify the order of the model which is to be fitted, a task which may not be straightforward when the output is not related to the input by a physical model of known form.

Interpreting the significance of the coefficients within the matrices which make up the state-space model may not be straightforward. A range of matrix algebra techniques can be used to manipulate the model to facilitate this interpretation. The most commonly used is a diagonalisation of the matrix A [Ogata, 1967]. This serves to remove interactions between states as time progresses, and in doing so reduces the system to a series of uncoupled first or second order systems, the coefficients of which can be interpreted as time constants or resonant frequencies respectively. However the application of such diagonalisations also has the effect of transforming the state variables which become linear combinations of the originals. Thus although the coefficient matrix can now be interpreted directly in terms of a series of isolated first (or in the case of complex eigenvalues, second) order systems, the physical significance of the state variables ceases to be readily apparent [Owens, 1981].

3.1.2 Auto-regressive/moving average (ARMA) models

This class of models seeks to represent the output of system in terms of a weighted sum of past inputs and outputs [Chatfield, 1976]:

$$y[t] = \sum_{k=1}^{oord} a_k y[t-k] + \sum_{k=0}^{iord} b_k x[t-k]$$

where:

$iord$ is the maximum delay associated with inputs to the model,

$oord$ is the maximum delay associated with the autoregressive section of the model, and

$\{a_k\}$ and $\{b_k\}$ are the model parameters.

The autoregressive model can be cast in a state-space representation, in which the system states are the past inputs to and outputs from the system. In this case the matrices of the state space

representation can become very large, although they will also be sparse. The practical benefits of such a representation are not immediately apparent.

A range of techniques are available for fitting models of this type [Box and Jenkins, 1976]. These methods are not, however, generally recursive, implying that data must be treated in batch form rather than sequentially. For the application considered here this is not a serious limitation, as complete sets of simulation and measured data will be available at the start of the analysis.

More serious is the requirement that most of the methods available, like those for fitting the state-space model, require the user to specify in advance the order of the model, in terms of the parameters i_{ord} and o_{ord} . A common approach to this task is to fit models of progressively higher orders until little further improvement in the quality of modelling is observed. This approach has the obvious disadvantage of greatly increasing the amount of computational effort required for the fitting process.

Whilst the moving average coefficients can be interpreted in terms of the contribution to the output from each of the model inputs, the autoregressive terms are not so amenable to analysis. Once again, however, techniques exist for the reduction of these coefficients to system poles and zeros, which have direct interpretation in terms of time constants and stability [Papoulis, 1977].

3.1.3 Transfer function modelling

The transfer function operator most commonly used to represent discrete time systems is the z operator. This operator, which represents a one time step delay, is defined [Jackson, 1986] by the relationship:

$$z^{-1} x[t] = x[t-1]$$

The z transform of a sequence is defined by the sum:

$$X[z] = \sum_{t=0}^{\infty} x[t] z^{-t}$$

The z transform has the useful property that the transform of the output of a system is found simply by multiplying the transform of the input by the system representation:

$$Y[z] = H[z] X[z]$$

The z transform of a system is a function of z, and can be represented in a variety of ways.

One common way is as the ratio of two polynomials in z, for example,

$$H[z] = \frac{\sum_{k=0}^{iord} b_k z^{-k}}{1 - \sum_{k=1}^{oord} a_k z^{-k}}$$

Expanding the above equation in terms of the definition of the z operator reveals that it is exactly equivalent to the equation used in Section 3.1.2 of this chapter to define the autoregressive/moving average model.

The principal advantage of using the z transform is that the transform of a system output is obtained simply by multiplying the input transform by the system transfer function. When these are of relatively simple form this can represent a significant simplification. However, when these are measured quantities the transforms will take the form of arbitrary polynomials of very high order, and computational savings may not be achieved.

3.1.4 Frequency response modelling

The discrete Fourier transform of the sequence $\{x[t]\}$ is defined [Papoulis, 1977] by:

$$X[\omega] = \sum_{t=0}^n x[t] e^{j\omega t \frac{2\pi}{n+1}}$$

where:

ω represents frequency and adopts integer values from 0 to n , and

$$j = \sqrt{-1}$$

Like the z transform, the Fourier transform has the property that the effect of a system on a given input can be determined simply by multiplying the transform of the input sequence by the transfer function of the system:

$$Y[\omega] = H[\omega] X[\omega]$$

However, this representation of the system in the frequency domain has the advantage over the z transform that the (generally complex) coefficients of the model $\{H[\omega]\}$, have a direct physical interpretation. At any given frequency ω the magnitude of $H[\omega]$ gives the system gain at that frequency, and the angle of $H[\omega]$ the phase shift.

Determining $\{H[\omega]\}$ is relatively straightforward once the observed input and output sequences have been transformed, and that process is aided greatly by the existence of the Fast Fourier Transform [Brigham, 1974]. Since both sequences will normally be real, a single transformation operation can be used to evaluate both the required transforms, a further economy in computation.

However, the resulting model, $\{H[\omega]\}$, will contain a very large number of parameters, typically $2(n+1)$, where n is the length of the input and output sequences (the 2 arises from the fact that the values of $H[\omega]$ are complex numbers).

3.1.5 Impulse response models

The impulse response model [Jong 1982] expresses the system output in terms of a weighted sum of past inputs alone:

$$y[t] = \sum_{k=0}^{\infty} h[k] x[t-k]$$

where:

the sequence $\{h[k]\}$ is the system impulse response.

The right hand side of this equation is known as the discrete convolution sum, and may also be written $h[k]*x[t]$.

The impulse response model can be viewed as a special case of the ARMA model, in which there are no autoregressive terms, and which has a moving average section of (potentially) infinite order. It is this potentially infinite order which makes it the most general representation of a linear, time invariant dynamic system: any such system can be completely described in terms of its impulse response, the set of coefficients $\{h[k]\}$.

The impulse response has an alternative interpretation in that it describes the output of the system when its input is an impulse, which is unity at time zero and zero at all other times. This yields a direct physical interpretation of the model coefficients in terms of the character of the system that they represent. The ability of the impulse response model to characterise any system completely stems from the fact that an arbitrary input sequence can be broken down into a series of scaled and time-shifted impulses. Because the system is postulated to be linear, the output due to a scaled impulse is simply the impulse response suitably scaled. Since the system is also time invariant, the response to an input shifted in time is simply the response shifted in time. Again because the system is linear the total output is simply the sum of the outputs due to each part of the decomposed input sequence.

A range of techniques exist for identifying the impulse response [Robinson, 1967] directly from records of input and output data. The impulse response of a system can also be generated from any of the models described above by applying an impulse to the identified model and calculating the output. In the case the z transform and frequency response models described above this process is particularly straightforward, as the both transforms evaluate to unity for the impulse input described. Therefore the impulse response of each model is simply the inverse transform of the system transfer function, which is simple to compute. As well as being a powerful modelling technique in its own right, the impulse response therefore represents a powerful tool for the interpretation of the results of fitting other models. If, however, the response is derived in this way, by first fitting an alternative model and then determining the impulse response of the fitted model, any assumptions or constraints implicit in the original model will be enforced on the resulting impulse response. For example, if the performance of a system of order say three is summarised by a first order auto-regressive system the resulting impulse response will be first order only, information about the higher order responses having been discarded when the auto-regressive model was fitted. For this reason it is clearly better to fit the impulse response directly to the data, rather than via some other model form.

3.2 Selection of an error model representation

The state-space model has been seen to provide a concise way of summarising the performance of a linear system in matrix form. When the model represents a given type of lumped physical system the coefficients of the matrix may be open to direct interpretation, but this will not generally be the case. Furthermore, the state-space notation requires that the order of the system be known before modelling commences, in order to set the dimension of the state vector and the system matrices. Auto-regressive/moving average models suffer from the same disadvantage: the number of both auto-regressive and moving average terms must be known in advance. Transfer function modelling (carried out using the z-transform for the discrete data considered here) suffers from the same shortcoming. In addition, the coefficients of either model are unlikely to be directly amenable to interpretation. Frequency response modelling permits interpretation in terms of the gain and phase shift of the system at different frequencies,

but this may not be the most useful characterisation of building or simulation model performance.

The impulse response model, however, yields a high degree of generality, at the same time requiring a minimum of a-priori information. Taken together, these properties imply that it will enable the development of a tool for comparing time series which destroys a minimum amount of information, and it is therefore considered to be the most appropriate model for the task in hand. Solution of the multiple input case normally requires that the user specify the maximum length of the response function, but this information is generally available from existing information about the system. In addition, the only penalty for choosing too long an impulse response is in terms of computational effort, and a conservative choice will therefore allow the model to be used with confidence.

The price which is paid for the high degree of physical interpretation available from the impulse response model is a lack of economy in parameters. As defined, the impulse response actually comprises an infinite series, although as discussed above this will generally be truncated. However, in the case of systems which are 'stiff', that is they have time constants which differ by many orders of magnitude, this truncation may only be possible after many hundred terms. One example has been produced [Norton, 1986] of a model which can be represented by an autoregressive process with only two parameters, but which requires an impulse response representation of 750 terms. In applications where rapid evaluation of the identified model is required this lack of compactness may represent a problem, but it is the very mechanism which makes the impulse response model amenable to the direct interpretation which is required here.

CHAPTER 4

SOLUTION OF THE IMPULSE RESPONSE MODEL

In this chapter the problem of identifying the impulse response of a single input system is first explored. This yields some valuable insight into the problems which can arise in identifying impulse response models. The identification procedure developed is then extended to the multiple input case which will ultimately be used to identify the sources of simulation errors. The implementation of the solution technique is described, together with a series of analytical tests of the solution software. These demonstrate that the solution and its implementation have been correctly derived, and also provide some interesting insight into the design of effective experiments for response function identification.

4.1 Single input case

For the case of a system with only a single input the impulse response model is as defined in Section 3.1.5 of Chapter 3:

$$y[t] = \sum_{k=0}^{\infty} h[k] x[t-k]$$

Here, the impulse response of the system has been assumed to be zero for times before $t=0$, ensuring the causality required of a physical system.

The equation allows the response of a system with a known impulse response to be generated for any arbitrary input. In approaching the problem of identifying the model, however, the output of the system is known, as is the input. The impulse response connecting the two data sequences is the required quantity, and the convolution equation given above must be inverted to obtain it, a process known as deconvolution [Cuénod and Durling, 1969].

If the input to the system is assumed to be zero before time $t=0$, that is that the system is quiescent before the start of the experiment, then the system equation can be inverted directly [Jong, 1982] to obtain a recursive solution for the impulse response $\{h[k]\}$:

$$h[0] = \frac{y[0]}{x[0]} \quad \text{followed by} \quad h[k] = \frac{y[k] - \sum_{i=0}^{k-1} h[i] x[k-i]}{x[0]}$$

This solution further requires that the input at time $t=0$ be non-zero. In practice this is a trivial requirement: if it is not initially satisfied it can be ensured by a simple translation of the time variable.

However, this does not represent a good way of estimating the system impulse response from recorded sequences of input and output data. Examining the solution reveals that the value of $h[0]$ is based on only one pair of data points. Any errors in $h[0]$, caused for example, by noise in those readings, will then be propagated through to the estimate of $h[1]$. Uncertainties in $h[0]$ and $h[1]$ are then carried forward into the estimate of $h[2]$ and so on. The process thus has very poor performance when faced with any noise in the measured data.

A second problem arises with the state of the system at time zero. The formulation of the solution above required the assumption that the inputs to the system were zero before $t=0$. In any real experiment there are likely to be driving forces (for example climate) which cannot be switched off up to the point at which the experiment begins. Thus at the start of the test the system retains some memory of the disturbances which occurred earlier. This effect is most pronounced in the measurements taken at the start of the experiment, the very values which are to be used to initiate the solution process by deducing $h[0]$.

The solution to both of these problems lies in averaging. The most simplistic approach is to make several identifications using the technique described above, and to then average the resulting impulse response estimates [Sage and Melsa, 1971]. An alternative technique, which guarantees the optimal use of all the data available in a single, continuous data set, is to formulate the convolution equation for the whole data set. This is most conveniently done using matrix notation, but it requires that a finite length is first assumed for the impulse response. Here, for the sake of brevity, we will assume that the impulse response contains only three non

zero terms, $h[0]$, $h[1]$ and $h[2]$, although in practice a greater length will generally be assumed.

The convolution equation can then be written:

$$\underline{y} = X \underline{h}$$

where:

$$\underline{y} = \begin{bmatrix} y[0] \\ y[1] \\ \vdots \\ y[n] \end{bmatrix}$$

$$X = \begin{bmatrix} x[0] & 0 & 0 \\ x[1] & x[0] & 0 \\ x[2] & x[1] & x[0] \\ x[3] & x[2] & x[1] \\ \vdots & \vdots & \vdots \\ x[n] & x[n-1] & x[n-2] \end{bmatrix}$$

and:

$$\underline{h} = \begin{bmatrix} h[0] \\ h[1] \\ h[2] \end{bmatrix}$$

This equation can be solved for \underline{h} , the required impulse response terms, using the pseudo-inverse of the non-square matrix X [Dorny, 1980]:

$$\underline{h} = (X^T X)^{-1} X^T \underline{y}$$

Carrying out the matrix transposition and multiplications yields:

$$X^T X = \begin{bmatrix} \sum_{t=0}^n x[t]x[t] & \sum_{t=1}^n x[t]x[t-1] & \sum_{t=2}^n x[t]x[t-2] \\ \sum_{t=1}^n x[t]x[t-1] & \sum_{t=0}^{n-1} x[t]x[t] & \sum_{t=1}^{n-1} x[t]x[t-1] \\ \sum_{t=2}^n x[t]x[t-2] & \sum_{t=1}^{n-1} x[t]x[t-1] & \sum_{t=0}^{n-2} x[t]x[t] \end{bmatrix}$$

a matrix which is symmetrical and positive definite, and:

$$X^T y = \begin{bmatrix} \sum_{t=0}^n x[t]y[t] \\ \sum_{t=1}^n x[t-1]y[t] \\ \sum_{t=2}^n x[t-2]y[t] \end{bmatrix}$$

This approach avoids the problems outlined earlier. The terms in the matrices are the result of totalling over the whole data set, and thus the effect of noise in the data is greatly reduced. If the experiment is significantly longer than the time constants of the system under investigation the influence of disturbances before the start of the experiment will also be minimal. The technique thus provides a robust way of obtaining impulse responses from streams of input and output variables.

The numerical requirements associated with this solution can, however, be taxing. There are two particular problem areas. The first is that the matrix $X^T X$ may be very close to singular, making its inversion prone to numerical problems. The physical reason underlying this is clear: adjacent columns of this matrix consist of observations of the same process shifted by one sampling interval. If the process is relatively slow moving in relation to the sampling interval this shift will have little effect, and adjacent columns will be closely related, leading to the

conditioning problem described.

This problem can be controlled by careful choice of matrix inversion algorithm. With care it may also be possible to exploit the symmetry and positive definiteness of the matrix. One example is the use of Choleski factorisation [Forsythe and Moler, 1967]. A matrix L is determined such that L is a lower triangular square root of $X^T X$, that is:

$$LL^T = X^T X$$

The positive definiteness of $X^T X$ guarantees the existence of L . Because of its lower triangular form, the inversion of L is trivial, and the required inverse of $X^T X$ can be quickly obtained from:

$$(X^T X)^{-1} = L^{-T} L^{-1}$$

The numerical advantages obtained from this technique are considerable. The accuracy obtained with this type of method is typically twice that available from methods not employing square roots. Other matrix inversion techniques can provide similar benefits. Examples are the Golub-Householder technique [Golub, 1965], or singular value decomposition techniques [Stoer and Bulirsch, 1980]. The latter has the advantage of explicitly diagnosing the source of numerical difficulties, but also the disadvantage of requiring two sets of matrix storage.

The second problem is simply the size of the equation set. Solving the equations for the $n+1$ terms of the impulse response requires the inversion of an $(n+1) \times (n+1)$ matrix. Depending on the system, this task may be impractical, an example being the system described in Section 3.2 which required 750 terms in its impulse response. One solution to this problem lies with recursive solution of the equation set, and such a solution has been found [Sage and Melsa, 1971]. However, it will not be exploited here. For the present application the number of terms required in the impulse response is such that the storage requirements remain relatively modest.

In some situations, there may be advance knowledge about the form of the impulse response, and this information may be used to advantage in the solution process. For example, if the physical model underlying the process dictates that the impulse response contains only non-negative entries this information may be used to greatly improve the quality of estimation [Commenges, 1984]. In the application described here no such a-priori knowledge will generally be available, and its assumption sacrifices the general form of the impulse response representation which first made it so appealing. Such constrained estimators will not, therefore, be pursued.

For all these solution variants, the quality of the solution obtained can be established by using the measured input sequences and the derived impulse responses in the convolution equation to 'reconstitute' the observed output (in this case simulation error). Close agreement between the reconstituted output and that observed indicates that the postulated linear time-invariant model is representing physical reality well.

4.1.1 Deconvolution using cross-covariance functions

An alternative solution to the deconvolution problem employs the cross-covariance function. The cross-covariance between two sequences of numerical data, $\{x[t]\}$ and $\{y[t]\}$, is defined [Box and Jenkins, 1980] by:

$$w_{xy}[k, t] = E((x[t] - \bar{x})(y[t+k] - \bar{y}))$$

where:

E denotes the expectation (or averaging) operation over all possible realisations of the stochastic processes $\{x[t]\}$ and $\{y[t]\}$,

\bar{x} denotes the mean of the sequence $\{x[t]\}$, and

\bar{y} denotes the mean of the sequence $\{y[t]\}$, that is:

$$\bar{x} = E(x[j]) \quad \bar{y} = E(y[j])$$

The auto-covariance function of a sequence is simply the cross-covariance of the sequence with itself. For the special case of processes which are stationary (that is, their statistics are not a function of time), the cross-covariance and auto-covariance functions are not a function of t . In this case, substituting the definitions of the covariance functions into the convolution equation, yields, with a little algebraic manipulation [Norton, 1986]:

$$r_{yx}[t] = \sum_{k=0}^{\infty} h[k] r_{xx}[t-k]$$

This rather remarkable result reveals that the cross-covariance between the input and output of a system is given by the convolution of the input signal auto-covariance with the system impulse response. The relation given is the discrete time form of the Wiener-Hopf equation [Norton, 1986].

The Wiener-Hopf equation can be directly inverted as before, to yield the impulse response of a system from the appropriate auto-covariance and cross-covariance functions. In this way, the computational advantages of direct solution can be retained, but the use of the cross-correlation function serves to produce the immunity to noise and initial conditions previously obtained by averaging.

As might be expected, the two solution techniques are very closely related. Indeed, the individual elements of the matrices $X^T X$ and $X^T y$ correspond directly to the values of the auto-correlation and cross-correlation functions. However, the derivation of the cross-correlation form given above has required that the input and output sequences be stationary. For the application considered here, this is a significant restriction: some of the system inputs are meteorological variables, and it has in the past been shown that these cannot be considered stationary [Irving, 1989]. The previous derivation, leading to the equivalent normal equations, did not require this assumption at any stage.

The resolution of this apparent contradiction lies with the definition of the covariance function. Conventionally this is defined in terms of expected values [Bendat and Piersol, 1986]. In this case the expectation is an ensemble average, that is it is assumed to be taken over all possible realisations of the random processes considered. However, when the correlation functions are actually estimated this is done by calculating mean values over a single sequence, and in this case, stationarity is required to ensure that the expected values, and hence cross-correlations, derived from one sample are representative of other samples taken at other times. The analysis presented here, however, seeks only to find the best set of coefficients from a single given dataset. In this situation the expectations represent average values across the dataset, and the assumption of stationarity is not required.

4.2 Multiple input case

The problem of identifying the sources of simulation errors requires that the relative contributions from a number of inputs, or driving forces, $\{x_1[t]\}$, $\{x_2[t]\}$, $\{x_3[t]\}$ etc to a single output be determined. Expressed in terms of the impulse responses to each of those inputs, $\{h_1[t]\}$, $\{h_2[t]\}$, $\{h_3[t]\}$ etc, the analysis model is:

$$y[t] = \sum_{k=0}^{\infty} h_1[k] x_1[t-k] + \sum_{k=0}^{\infty} h_2[k] x_2[t-k] + \dots + \sum_{k=0}^{\infty} h_m[k] x_m[t-k]$$

where:

m is the number of inputs to the model.

For the purposes of solution, this equation can again be cast in matrix form. Here, for the sake of brevity, we consider a case with two independent variables, the impulse response of each of which has only three non-zero elements. The extension to larger cases (for example the five independent variables with impulse responses of 49 elements each of which will ultimately be implemented) is straightforward. Expanding the impulse response model for this case gives:

$$\underline{y} = X \underline{h}$$

where:

$$\underline{y} = \begin{bmatrix} y[0] \\ y[1] \\ \vdots \\ y[n] \end{bmatrix}$$

$$X = \begin{bmatrix} x_1[0] & 0 & 0 & x_2[0] & 0 & 0 \\ x_1[1] & x_1[0] & 0 & x_2[1] & x_2[0] & 0 \\ x_1[2] & x_1[1] & x_1[0] & x_2[2] & x_2[1] & x_2[0] \\ x_1[3] & x_1[2] & x_1[1] & x_2[3] & x_2[2] & x_2[1] \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1[n] & x_1[n-1] & x_1[n-2] & x_2[n] & x_2[n-1] & x_2[n-2] \end{bmatrix}$$

and:

$$\underline{h} = \begin{bmatrix} h_1[0] \\ h_1[1] \\ h_1[2] \\ h_2[0] \\ h_2[1] \\ h_2[2] \end{bmatrix}$$

This equation can be solved for \underline{h} , the required impulse response terms, using the pseudo-inverse of the non-square matrix X [Dorny, 1980]:

$$\underline{h} = (X^T X)^{-1} X^T \underline{y}$$

Carrying out the matrix transposition and multiplications yields:

$$X^T X =$$

$$\begin{bmatrix} \sum_{i=0}^n x_1[i]x_1[i] & \sum_{i=1}^n x_1[i]x_1[i-1] & \sum_{i=2}^n x_1[i]x_1[i-2] & \sum_{i=0}^n x_1[i]x_2[i] & \sum_{i=1}^n x_1[i]x_2[i-1] & \sum_{i=2}^n x_1[i]x_2[i-2] \\ \sum_{i=1}^n x_1[i]x_1[i-1] & \sum_{i=0}^{n-1} x_1[i]x_1[i] & \sum_{i=1}^{n-1} x_1[i]x_1[i-1] & \sum_{i=1}^n x_1[i-1]x_2[i] & \sum_{i=0}^{n-1} x_1[i]x_2[i] & \sum_{i=1}^{n-1} x_1[i]x_2[i-1] \\ \sum_{i=2}^n x_1[i]x_1[i-2] & \sum_{i=1}^{n-1} x_1[i]x_1[i-1] & \sum_{i=0}^{n-2} x_1[i]x_1[i] & \sum_{i=2}^n x_1[i-2]x_2[i] & \sum_{i=1}^{n-1} x_1[i-1]x_2[i] & \sum_{i=0}^{n-2} x_1[i]x_2[i] \\ \sum_{i=0}^n x_1[i]x_2[i] & \sum_{i=1}^n x_1[i-1]x_2[i] & \sum_{i=2}^n x_1[i-2]x_2[i] & \sum_{i=0}^n x_2[i]x_2[i] & \sum_{i=1}^n x_2[i]x_2[i-1] & \sum_{i=2}^n x_2[i]x_2[i-2] \\ \sum_{i=1}^n x_1[i]x_2[i-1] & \sum_{i=0}^{n-1} x_1[i]x_2[i] & \sum_{i=1}^{n-1} x_1[i-1]x_2[i] & \sum_{i=1}^n x_2[i]x_2[i-1] & \sum_{i=0}^{n-1} x_2[i]x_2[i] & \sum_{i=1}^{n-1} x_2[i]x_2[i-1] \\ \sum_{i=2}^n x_1[i]x_2[i-2] & \sum_{i=1}^{n-1} x_1[i]x_2[i-1] & \sum_{i=0}^{n-2} x_1[i]x_2[i] & \sum_{i=2}^n x_2[i]x_2[i-2] & \sum_{i=1}^{n-1} x_2[i]x_2[i-1] & \sum_{i=0}^{n-2} x_2[i]x_2[i] \end{bmatrix}$$

and:

$$X^T \underline{y} = \begin{bmatrix} \sum_{i=0}^n x_1[i]y[i] \\ \sum_{i=1}^n x_1[i-1]y[i] \\ \sum_{i=2}^n x_1[i-2]y[i] \\ \sum_{i=0}^n x_2[i]y[i] \\ \sum_{i=1}^n x_2[i-1]y[i] \\ \sum_{i=2}^n x_2[i-2]y[i] \end{bmatrix}$$

The elements of the matrices $X^T X$ and $X^T \underline{y}$ again correspond to the cross-correlations and auto-correlations used in the time domain solution. Now, of course, the input auto-correlation function includes the cross-correlations between separate inputs, as well as the auto-correlation functions of each input.

Once again, the observed output sequence may be reconstituted using the observed inputs and the estimated impulse response terms. Again, comparison between the reconstituted and observed outputs gives a measure of how well the fitted model represents the observed data, but it now has a further rôle. The contribution to the single output from each of the inputs can now be assessed, providing a direct measure of the relative importance of each input.

Clearly the analysis required for the multivariate case is much more complex than that for the univariate case, because of the possibility of cross-correlations between the various input sequences. When these inter-correlations do not exist, or can be ignored, the multivariate solution can be replaced by a univariate solution for each impulse response in turn. In this case the reconstitution of the errors using the model equation provides a test of how well the model derived using the simplified analysis accounts for the errors observed.

4.3 Propagation of mean values through the deconvolution process

In all of the preceding analysis the input and output sequences have been processed directly, in their units of measurement. The combination of choice of measurement units and nature of the driving force may result in the measured values being relatively small variations about a mean value. In computing quantities of the form

$$\sum_{t=k}^n x[t] x[t-k]$$

the presence of the relatively large mean value may cause a significant reduction in numerical precision. If the sequence $\{x[k]\}$ has mean value \bar{x} then a corresponding zero mean sequence can be defined by:

$$x'[t] = x[t] - \bar{x}$$

Simple substitution reveals that:

$$\sum_{t=k}^n x[t] x[t-k] = \sum_{t=k}^n x'[t] x'[t-k] + (n-k) \bar{x}^2$$

Depending on the quantity which x represents, the $(n-k)x^2$ term may become very large. For example, for hourly measurements of solar radiation over a 50 day experimental period, it may be in excess of 50 000 000. It is undesirable to carry such a large offset through the analysis process, as it will reduce the degree of precision available. If:

$$y[t] = h[k] * x[t]$$

then the linearity of the convolution sum means that:

$$y[t] = y'[t] + \bar{y} = h[t] * x'[t] + h[t] * \bar{x}$$

Substitution in the convolution equation reveals that:

$$h[k] * \bar{x} = \bar{y}$$

and therefore:

$$y'[t] = h[k] * x'[t]$$

In other words, if a pair of input and output sequences $\{x[t]\}$ and $\{y[t]\}$ are linked by the impulse response $\{h[t]\}$ then the corresponding zero mean sequences $\{x'[t]\}$ and $\{y'[t]\}$ will be linked by the same impulse response. This implies that all input and output sequences can be reduced to zero mean form before analysis without any effect on the estimated $\{h[t]\}$, avoiding the numerical problems referred to above.

In all the applications of the impulse response fitting method which follow, all sequences are

first reduced to zero mean. The resulting impulse responses are then used to reconstitute the output, using the input data directly, in its non-zero mean form. It is possible that the reconstituted output may not then have the correct mean value. This is because information about the mean values of all the sequences has been discarded during the analysis process. The problem can be resolved by modifying the analysis model to:

$$y[t] = \sum_{k=0}^{\infty} h_1[k] x_1[t-k] + \sum_{k=0}^{\infty} h_2[k] x_2[t-k] + \dots + \sum_{k=0}^{\infty} h_m[k] x_m[t-k] + c$$

where:

c is a constant.

If required, the value of c can be obtained by comparing the mean of the measured output data with the mean of the reconstituted output.

4.4 Implementation

It is clear from the derivation of Section 4.2 that the computational requirements of the method are determined by the number of input variables, the number of data points, and the maximum delay allowed in the resulting impulse responses. In general, the storage requirements will be dominated by the matrix $\mathbf{X}^T\mathbf{X}$, and subsequently, its inverse. We denote the dimension of these matrices as $p \times p$ where $p = (\text{number of input variables}) \times (\text{maximum delay} + 1)$. The storage requirement is therefore of the order p^2 .

The computational effort required is dominated by two processes: determining the elements of $\mathbf{X}^T\mathbf{X}$, and inverting it. The former process requires that a total of $(p^2+p)/2$ different sums be generated for each input variable. Each sum requires (number of data points) multiplications, and thus the total number of multiplications required is $(p^2+p)/2 \times (\text{number of data points})$. The inversion of the resulting matrix requires of the order p^3 multiplications.

A computer program has been prepared which carries out the analysis described in Section 4.2. The program is written in Turbo Pascal, version 6.0, and runs on an IBM Personal Computer or compatible. The program is modular, which allows rapid modification, for example the installation of alternative equation solving routines.

The program reads a series of data records, each of which contains the input variables and corresponding output sequences for one time step. All data sequences are transformed to zero-mean, and the input auto-correlation matrix ($X^T X$) is generated. Next the input/output cross-correlation vectors ($X^T y$) are generated for each of the output streams being processed. The resulting linear equation set is then solved. Currently three matrix inversion schemes are offered. The first two of these are standard published routines and do not exploit the symmetry and positive definiteness of the matrix equation being solved:

- a Gauss-Jordan method with full pivoting to maintain numerical stability and avoid unnecessary loss of numerical precision [Press, Flannery, Teulosky and Vetterling, 1989]. The solution vectors are generated by simultaneously carrying out the row operations on each of the input/output cross-correlation vectors. This approach yields a much higher precision than the alternative of using the matrix inverse to generate the solutions by matrix multiplication,
- a general decomposition routine, which uses Crout's algorithm to break the matrix into upper and lower triangular factors. These factor matrices can then be used to generate the solution vector [Press, Flannery, Teulosky and Vetterling, 1989]. Once again, pivoting, this time partial, is used to ensure numerical stability. Because the actual inverse of the matrix is not required in this application, this algorithm allows a substantial saving in computational effort over the Gauss-Jordan scheme.

The third option is a routine specially written for this task which exploits the special form of the matrix to the full:

- a Choleski factorisation method which decomposes the matrix into square roots. As well as giving the improvement in numerical precision described in Section 4.1, this algorithm

gives a considerable further speed increase.

The derived impulse responses are then used, in conjunction with the original input data, to reconstitute the output data streams. The overall reconstituted output and the individual contributions from each input are then written to a file.

The software was originally developed on a 12 MHz 286-based PC. With a maths coprocessor the program took typically 8 hours to derive 48 hour long impulse responses from a data set containing 1200 records each containing 5 outputs and 7 inputs when using the Choleski factorisation solution routine. On a 66 MHz 486DX machine this falls to less than 1 hour, and it is assumed that as PC technology advances this figure will continue to fall.

4.5 Verification of the solution process

Clearly the derivation and implementation of the multiple input solution to the impulse response model are complex processes, and may be subject to error. In this section a simple test of the solution process is described. As well as verifying that the solution can be correctly obtained, the exercise provides some useful insight into how the identification process should be used.

The testing process consists of first generating a model output, using known input sequences and a set of known impulse responses. The solution software is then used to obtain the impulse responses from the input and output sequences, as it will be in real applications. These can be compared with the known values from which the test data was generated, to confirm that the system is functioning correctly.

The input sequences and impulse response functions chosen for this process can, of course, be arbitrary. If a very large number of tests were to be conducted it could be argued that they should be generated randomly, to minimise the chances of a fortuitous choice masking a fault in the software. A different approach has been adopted here. In order to test the solution process under the most realistic of operating conditions, the input sequences have been derived

from real weather and room operation data. Real records of:

- 1) pseudo random heater operation
- 2) external air temperature,
- 3) solar radiation,
- 4) wind velocity, and
- 5) sky temperature

were obtained. Each sequence was scaled to have a mean of unity. This scaling operation is arbitrary, and it serves only to ensure that impulse responses of comparable magnitudes produce comparable contributions to the system output. The reason for this choice of inputs was that it ensured that the method was tested using inputs which had auto-correlation and cross-correlation functions the same as those with which the method would eventually be used. The ability to deal with these correlations amongst the inputs is an important requirement of the method in the proposed application.

The impulse responses assumed for each input were chosen with less regard for realism, on the grounds that they should provide a thorough test of the method's ability to extract all kinds of response. Indeed, given that these represent the responses of simulation error to the inputs driving both model and reality it is hard to envisage what would constitute a 'realistic' set. Depending on the cause of the errors within the model the responses could take on almost any form. The response functions chosen for each of the five inputs listed above were:

- 1) a simple exponential decay, with a time constant of approximately $1\frac{1}{2}$ hours for the room heater,
- 2) an oscillatory decay, again with a time constant of $1\frac{1}{2}$ hours for external temperature,
- 3) a very slow ramp up, followed by a slow ramp back down for solar radiation. The total length of this response is 19 hours,
- 4) a pulse response, corresponding to a simple moving average of the wind speed input. This was chosen because the sharp edges of the pulse might be expected to excite any instability inherent in the solution software, and finally,
- 5) a zero response was attributed to the sky temperature input, meaning that this input made

no contribution to the output of the system. This was chosen to reveal whether the realistic correlations between the inputs could cause the software to give an erroneous indication that an input made a contribution to the output when that effect was in fact due to alternative input. This is clearly important in the proposed application, as such an erroneous indication would suggest that an input was contributing to simulation error when in fact it was not, causing incorrect conclusions to be drawn about the functioning of a model.

The impulse responses used are tabulated in Table 4.1.

Time delay (hours)	Response to input 1 (Heater)	Response to input 2 (Ext temp)	Response to input 3 (Solar)	Response to input 4 (Wind vel)	Response to input 5 (Sky temp)
0	1.000	1.000	0.000	1.000	0.000
1	0.500	-0.500	0.100	1.000	0.000
2	0.250	0.250	0.200	1.000	0.000
3	0.125	-0.125	0.300	1.000	0.000
4	0.063	0.063	0.400	1.000	0.000
5	0.031	-0.031	0.500	1.000	0.000
6	0.016	0.016	0.600	0.000	0.000
7	0.000	0.000	0.700	0.000	0.000
8	0.000	0.000	0.800	0.000	0.000
9	0.000	0.000	0.900	0.000	0.000
10	0.000	0.000	1.000	0.000	0.000
11	0.000	0.000	0.900	0.000	0.000
12	0.000	0.000	0.800	0.000	0.000
13	0.000	0.000	0.700	0.000	0.000
14	0.000	0.000	0.600	0.000	0.000
15	0.000	0.000	0.500	0.000	0.000
16	0.000	0.000	0.400	0.000	0.000
17	0.000	0.000	0.300	0.000	0.000
18	0.000	0.000	0.200	0.000	0.000
19	0.000	0.000	0.100	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000

Table 4.1: Assumed impulse responses for analytical test of solution process

A sequence of output values was then generated by convolving the input streams described earlier with the impulse responses of Table 4.1. Next, the computer program described in

Section 4.4 was used to recover the impulse responses of Table 4.1. Table 4.2 shows the results, and Figures 4.1 to 4.5 show the assumed and recovered impulse responses graphically.

Time delay (hours)	Response to input 1 (Heater)	Response to input 2 (Ext temp)	Response to input 3 (Solar)	Response to input 4 (Wind vel)	Response to input 5 (Skytemp)
0	1.007	0.996	0.071	0.974	0.002
1	0.506	-0.501	0.062	1.008	0.001
2	0.257	0.247	0.198	0.981	-0.000
3	0.132	-0.126	0.293	0.981	0.000
4	0.068	0.061	0.393	0.983	0.000
5	0.037	-0.034	0.528	0.979	0.000
6	0.021	0.014	0.603	0.014	0.000
7	0.005	-0.002	0.708	0.009	-0.000
8	0.005	-0.002	0.801	0.014	-0.000
9	0.005	-0.002	0.904	0.012	0.000
10	0.004	-0.001	0.979	0.005	-0.000
11	0.002	-0.002	0.886	0.000	-0.000
12	0.002	-0.001	0.820	-0.002	-0.000
13	0.002	0.000	0.677	0.001	-0.000
14	0.002	-0.000	0.598	-0.010	0.000
15	0.002	0.003	0.491	-0.003	-0.001
16	0.002	-0.001	0.402	-0.004	0.000
17	0.001	-0.000	0.296	0.0021	-0.001
18	0.003	-0.001	0.196	0.011	0.000
19	0.004	-0.001	0.112	-0.007	-0.000
20	0.005	-0.001	0.001	-0.007	-0.000
21	0.004	-0.002	0.006	0.012	-0.000
22	0.003	0.001	-0.012	0.009	-0.000
23	0.004	-0.001	-0.006	0.002	0.000
24	0.003	0.001	0.021	0.004	-0.001

Table 4.2: Derived impulse responses from test data set (maximum delay = 24 hours)

Table 4.3 shows the errors in these derived impulse responses.

Time delay (hours)	Response to input 1 (Heater)	Response to input 2 (Ext temp)	Response to input 3 (Solar)	Response to input 4 (Wind vel)	Response to input 5 (Skytemp)
0	0.007	-0.004	0.071	-0.026	0.002
1	0.006	-0.001	-0.038	0.008	0.001
2	0.007	-0.003	-0.002	-0.019	-0.000
3	0.007	-0.001	0.007	-0.019	0.000
4	0.005	-0.002	-0.007	-0.017	0.000
5	0.006	-0.003	0.028	-0.021	0.000
6	0.021	0.013	0.003	0.014	0.000
7	0.005	-0.002	0.008	0.009	-0.000
8	0.005	-0.002	0.001	0.014	-0.000
9	0.005	-0.002	0.004	0.012	0.000
10	0.004	-0.001	-0.021	0.005	-0.000
11	0.002	-0.002	-0.014	0.000	-0.000
12	0.002	-0.001	0.020	-0.002	-0.000
13	0.002	0.000	-0.023	0.001	-0.000
14	0.002	-0.000	-0.002	-0.010	0.000
15	0.002	0.003	-0.009	-0.003	-0.000
16	0.002	-0.001	0.002	-0.004	0.000
17	0.001	-0.000	-0.004	0.002	-0.001
18	0.003	-0.001	-0.004	0.011	0.000
19	0.004	-0.001	0.012	-0.007	-0.000
20	0.005	-0.001	0.001	-0.007	-0.000
21	0.004	-0.002	0.006	0.012	-0.000
22	0.003	0.001	-0.012	0.009	-0.000
23	0.004	-0.001	-0.006	0.002	0.000
24	0.003	0.000	0.021	0.004	-0.001

Table 4.3: Errors in derived impulse responses from test data set (maximum delay = 24 hours)

The errors given in Table 4.3 are summarised in terms of their mean and standard deviation in Table 4.4.

Time delay (hours)	Response to input 1 (Heater)	Response to input 2 (Ext temp)	Response to input 3 (Solar)	Response to input 4 (Wind vel)	Response to input 5 (Skytemp)
Mean error	0.005	-0.001	0.001	-0.001	0.000
Standard deviation	0.004	0.003	0.020	0.012	0.001

Table 4.4: Statistics of errors in derived impulse responses from test data set
(maximum delay = 24 hours)

These results demonstrate that the analysis described is indeed able to recover the impulse responses from Table 4.1 which were used to generate the test data set.

When designing an experiment to produce data for analysis it is clearly of interest to see the effect of using smaller numbers of data points, that is, conducting a shorter trial. The analysis described above (based on 1200 points) was thus repeated on subsets of the original data set, consisting of the first 600, the first 300 and the first 150 points. Table 4.5 shows standard deviation of the resulting estimates of the impulse responses. Figure 4.6 shows the effect of reducing the number of data points graphically.

Number of data points	Response to input 1 (Heater)	Response to input 2 (Ext temp)	Response to input 3 (Solar)	Response to input 4 (Wind vel)	Response to input 5 (Skytemp)
1200	0.004	0.003	0.020	0.012	0.001
600	0.031	0.059	0.686	0.123	0.017
300	0.059	0.122	1.026	0.374	0.031
150	0.252	0.192	1.333	0.880	0.059

Table 4.5: Standard deviations of errors in derived impulse responses from test data set as number of data points is decreased (maximum delay = 24 hours)

Tables 4.4 and 4.5 demonstrate clearly that the accuracy which can be obtained from the method is strongly influenced by the nature of the input to the system. Input stream 1, the assumed random heater input, and stream 5, night sky temperature, both allow very good identification from the full 1200 point dataset, and produce acceptable results from as few as 300 points. The response to input stream 3, however, which was derived from ambient temperature, proves less easy to determine even from the full dataset, and it is effectively impossible to recover it from the shorter sets. Inspection of the responses to external temperature actually recovered reveal that the large variances shown in Table 4.5 are the result of oscillation at the start of the estimated response. There is a simple physical interpretation of this problem. External temperature is a slow moving (or strongly auto-correlated) process, that is the value at any given hour is strongly dependent on the value at the previous hour. It thus fails to excite the more rapid responses of the system, and so it is not possible to obtain a good characterisation of that part of the impulse response.

The technique described for the solution of the multivariate impulse response model requires that the user specify a maximum length for the system impulse responses. In addition, the computational effort required for the identification depends critically on this assumption, and the question of the penalties associated with selecting too short a length therefore arises. Table 4.6 shows the recreated impulse responses when a maximum length of only 12 hours is assumed. It is clear from Table 4.1 that this assumption is inappropriate, as the response to

solar radiation is of length 19 hours. It is, nevertheless, an error which could easily occur in a real application of the technique, where the actual length of the impulse responses is unknown.

Time delay (hours)	Response to input 1 (Heater)	Response to input 2 (Ext temp)	Response to input 3 (Solar)	Response to input 4 (Wind vel)	Response to input 5 (Skytemp)
0	1.002	1.002	0.753	1.044	0.002
1	0.497	-0.525	-0.316	1.069	0.001
2	0.242	0.219	0.288	1.026	0.001
3	0.128	-0.080	-0.068	1.003	-0.002
4	0.059	0.074	0.322	0.977	-0.006
5	0.026	-0.054	0.431	0.921	0.005
6	0.017	0.052	0.527	-0.078	-0.001
7	-0.006	0.006	0.904	-0.024	-0.001
8	-0.014	-0.034	0.710	-0.144	0.002
9	-0.002	0.011	0.694	0.152	0.001
10	-0.006	-0.039	1.406	0.005	0.004
11	-0.001	0.066	-0.977	-0.089	-0.004
12	-0.005	-0.350	5.426	0.022	0.036

Table 4.6: Derived impulse responses from test data set (maximum delay = 12 hours)

Table 4.6 demonstrates the sort of erroneous conclusions which could be drawn if the maximum length assumed for the impulse responses is insufficient. Specifically:

- the estimated response to variable 3 (the response that exceeds the maximum assumed length) shows large, erroneous, entries in the tail of the response, and
- this disturbance also affects the estimates of the impulse responses to the other inputs (specifically variable 2), even though these responses themselves do not contain terms beyond the assumed maximum lengths.

These conclusions demonstrate clearly the dangers of assuming too short an impulse response

length. However, they do not indicate a failure of the solution technique, but rather of the way in which it has been applied. The method has been constrained to identify the best linear fit between the output sequence and the 13 most recent values of all the input variables. In reality, there is also a contribution to the output from the 14th most recent value of input 3 (and indeed the 15th, 16th etc). This contribution must be accounted for by the constrained analysis process in terms of the input variables which are available in the assumed model. The structure of the input data used is such that the 14th most recent input value is closely correlated with the 13th, and thus the solution process will attempt to account for the contribution from the 14th term by adjusting the coefficient of the 13th term. Thus the large, spurious, impulse response terms seen in Table 4.6 and on Figure 4.3 appear in the tail of the impulse response.

However, there are also correlations between the different input streams. Thus the solution method may be able to improve the overall fit of the model by also including further erroneous terms in the responses to other input streams. This can be seen in the estimated response to variable 2 in Table 4.6 and on Figure 4.2.

These explanations are further confirmed reinforced by the fact that the response to input 1, the randomised input which does not correlate with any of the other driving forces, emerges uncorrupted. This demonstrates a further significant advantage of, wherever possible, randomising the experimental driving forces.

4.6 Guidelines for the design of impulse response identification experiments

The experiences described in this chapter lead to a number of conclusions as to how best to design experiments for the identification of impulse response models:

- it is clear that the quality of response identification depends very strongly on the characteristics of the variables driving the test building. In particular, identification is much harder in the face of slowly moving, or strongly auto-correlated excitation. Obviously the nature of some driving forces (climate) is not under the control of the experimenter, but whenever possible driving forces should be chosen to have minimum auto-correlations.

Randomising a variable is one way of achieving this. Steady driving forces (for example a heater operated at constant power) should be avoided at all costs.

- cross-correlations between the forces driving the test building also reduce the quality of identification. Once again, in some cases these interactions will be unavoidable, but wherever possible inputs should be chosen not to correlate with each other, and not to correlate with other, predetermined, inputs. Again, randomisation is one way of achieving this.
- excessively large entries in the tail of a response may indicate that the identification process is not considering sufficiently long time delays. The identification should be repeated including more response terms.

FIGURES

Analytical test of impulse response model solution Response to test input stream 1

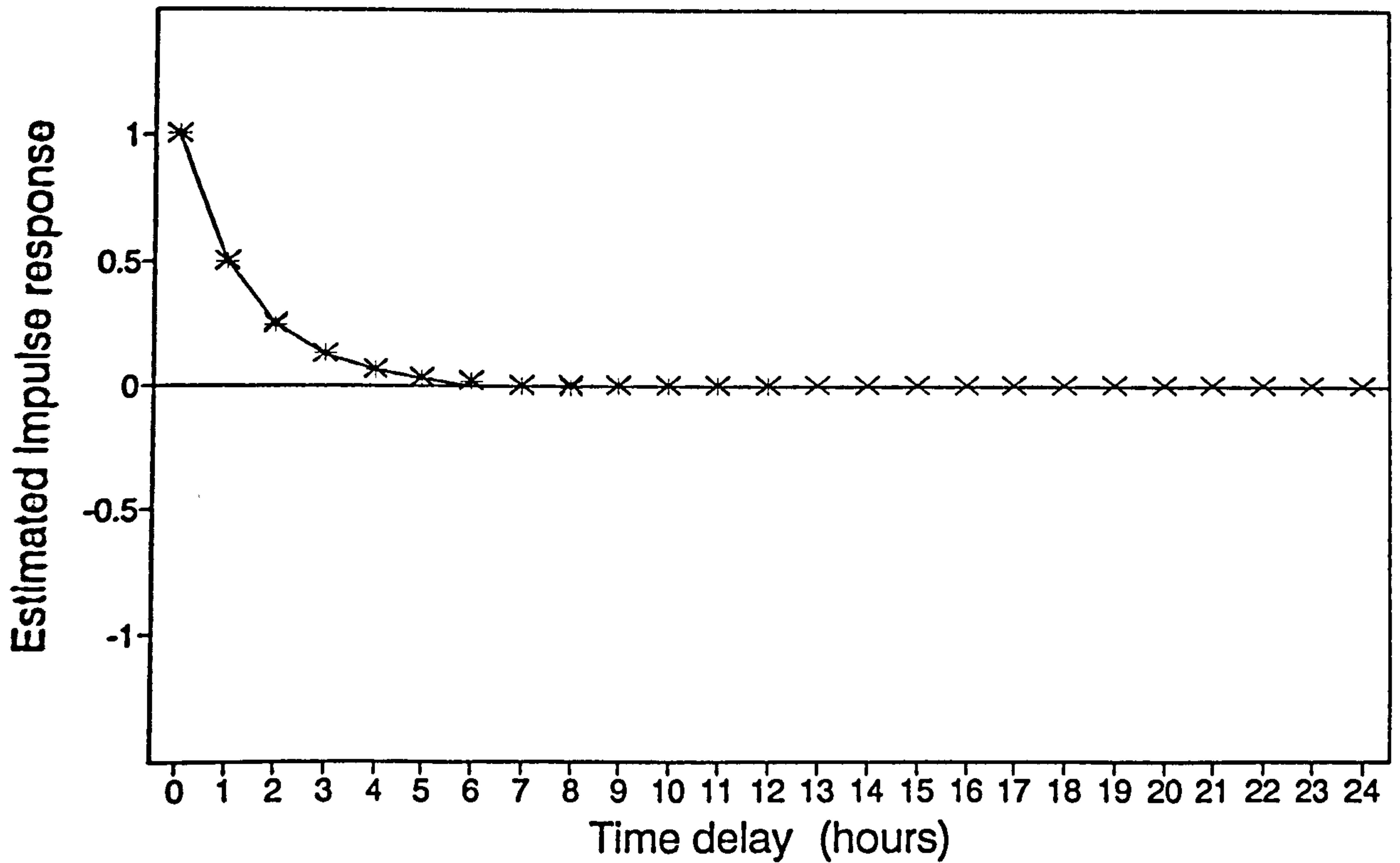
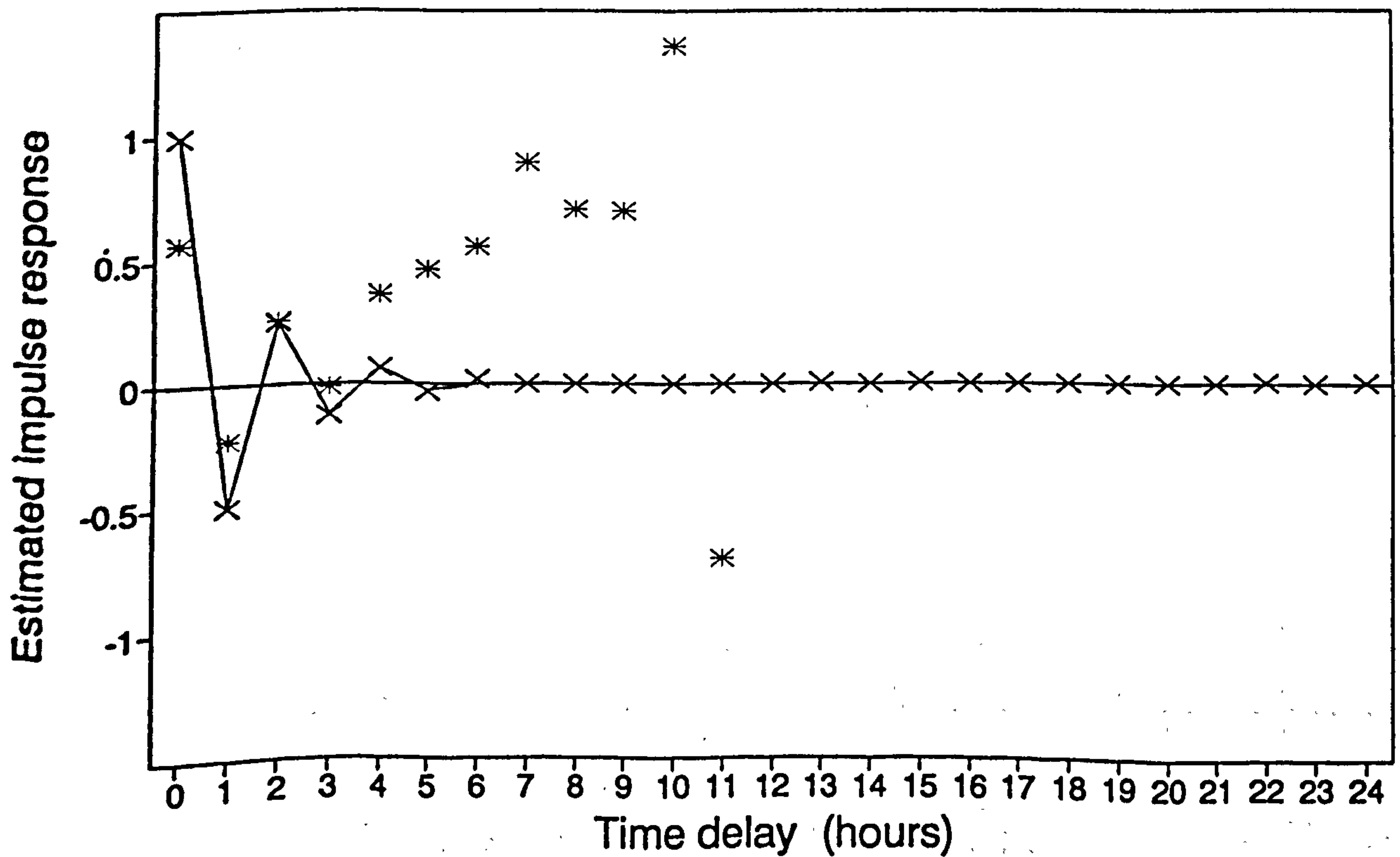


Figure 4.1

— actual response × identified (24 hr) * identified (12 hr)

Analytical test of impulse response model solution Response to test input stream 2



Analytical test of impulse response model solution Response to test input stream 3

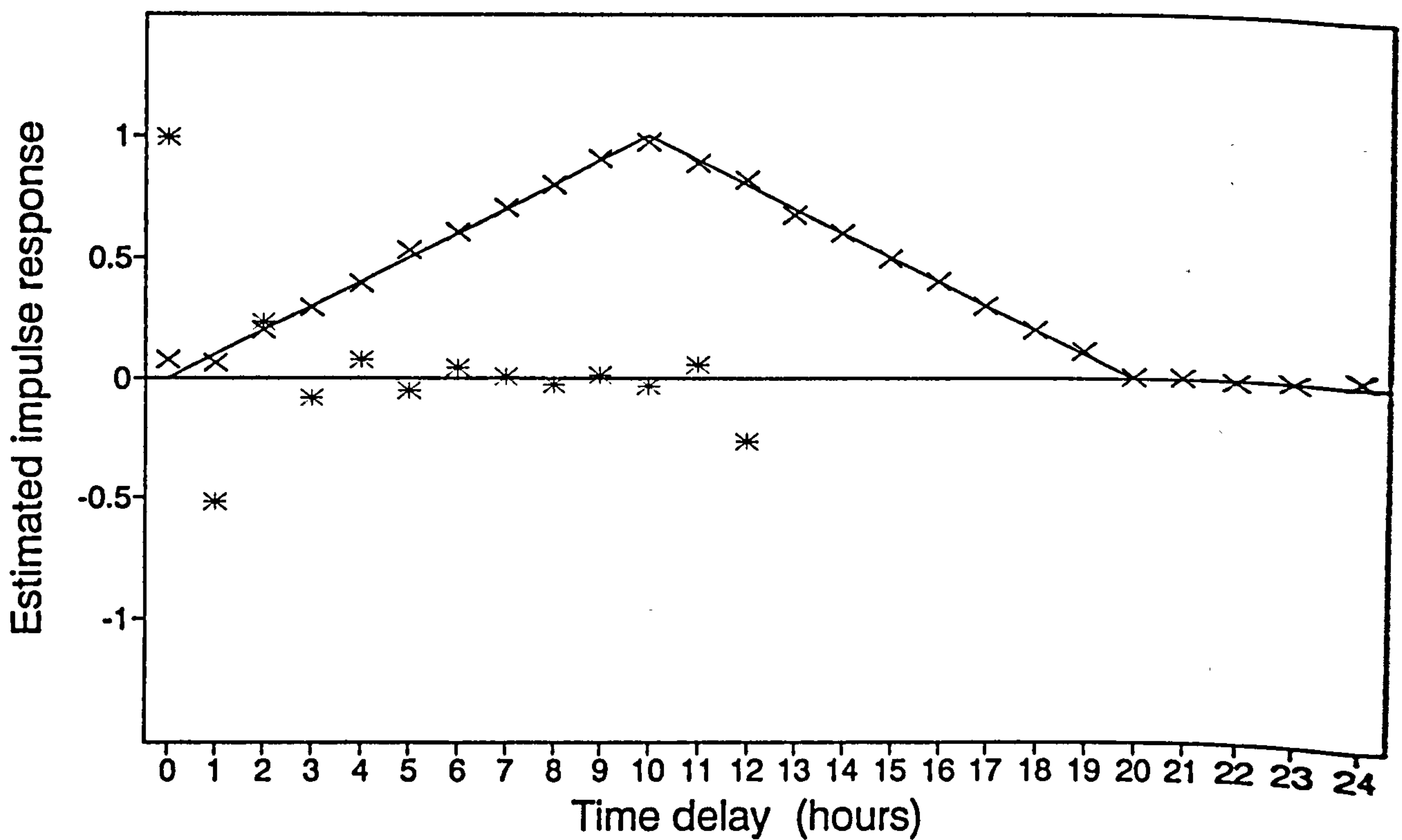


Figure 4.3

— actual response × identified (24 hr) * identified (12 hr)

Analytical test of impulse response model solution Response to test input stream 4

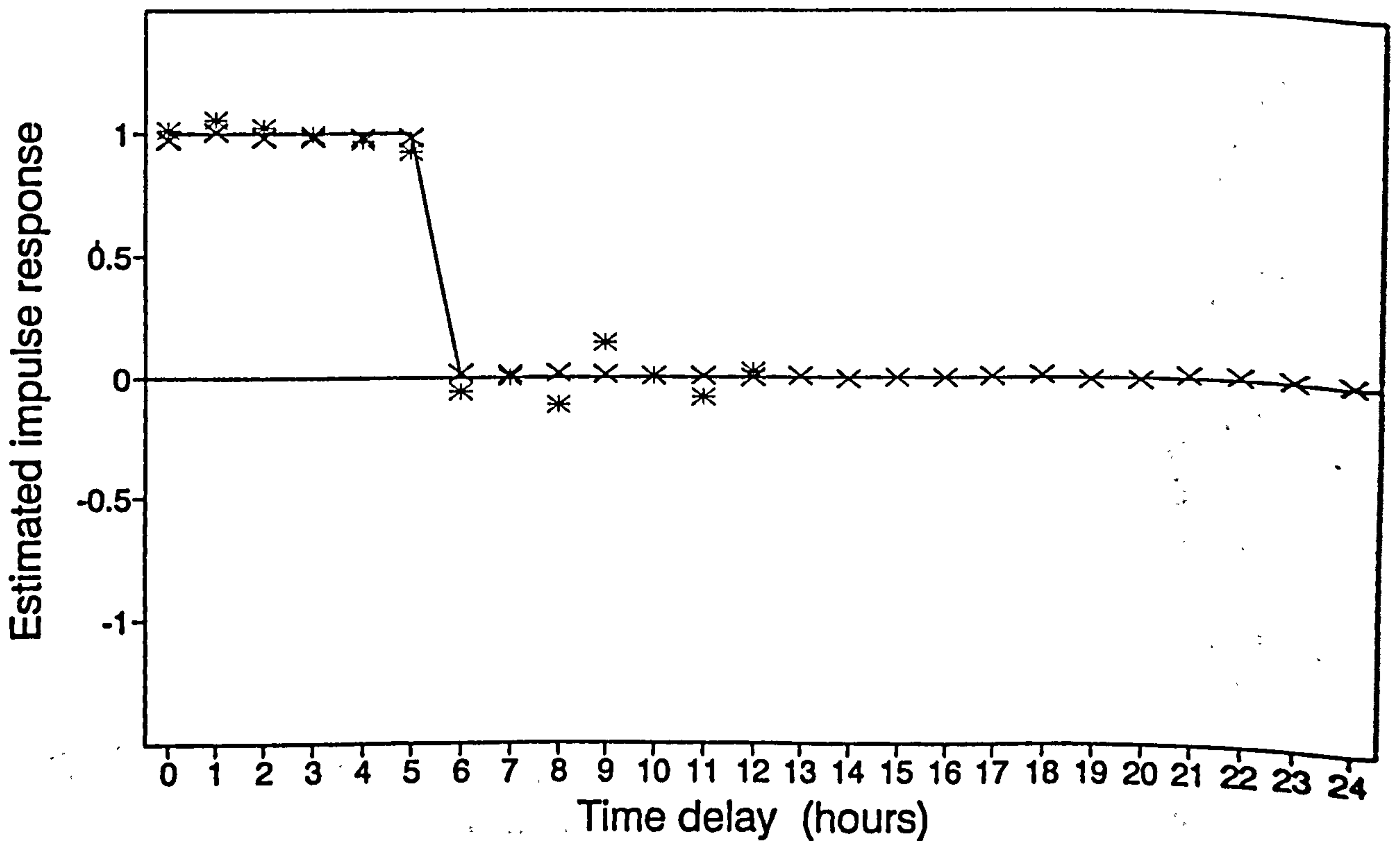


Figure 4.4

— actual response × identified (24 hr) * identified (12 hr)

Analytical test of impulse response model solution
Response to test input stream 5

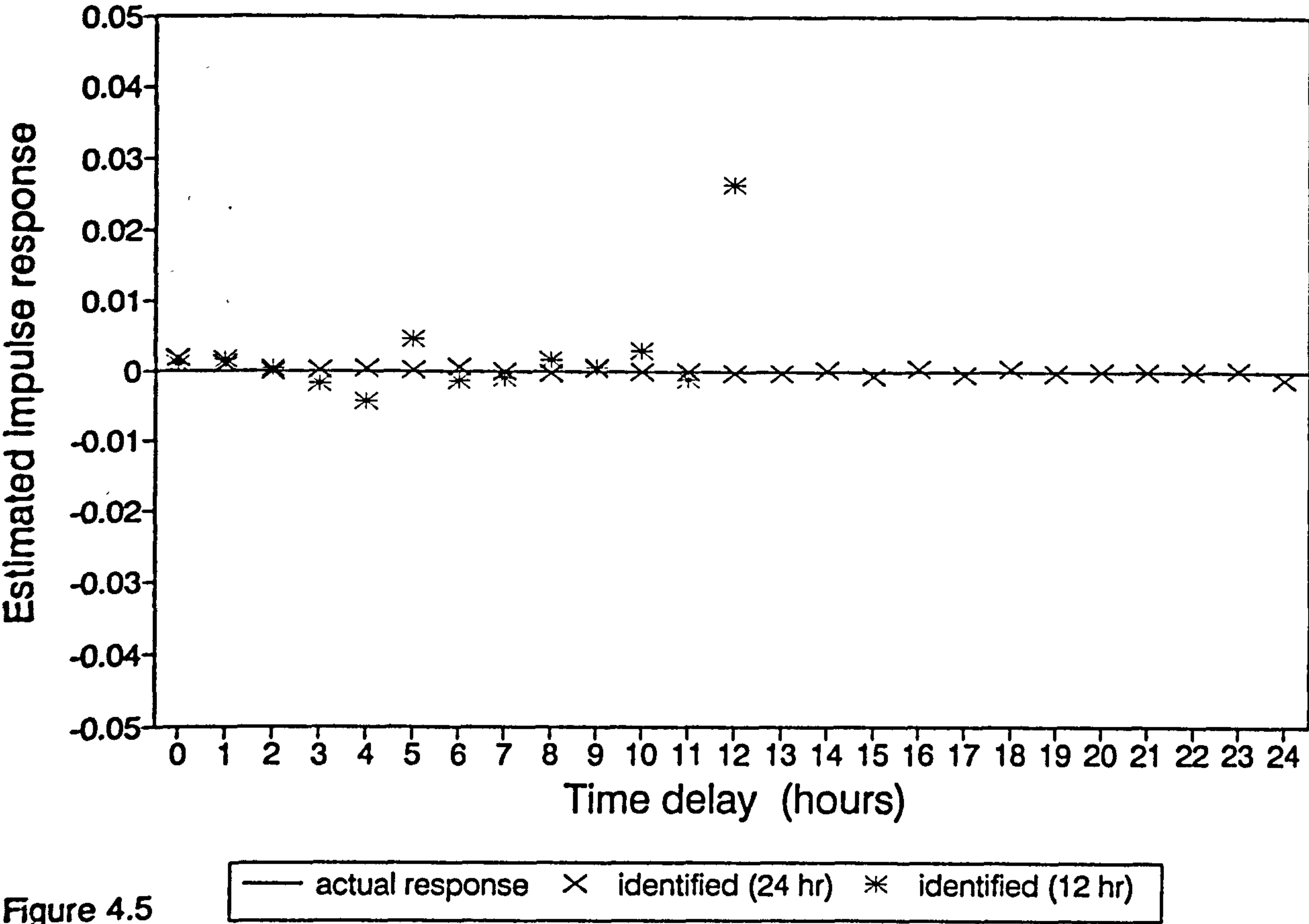


Figure 4.5

Analytical test of impulse response model solution
Variation of estimation error with number of points

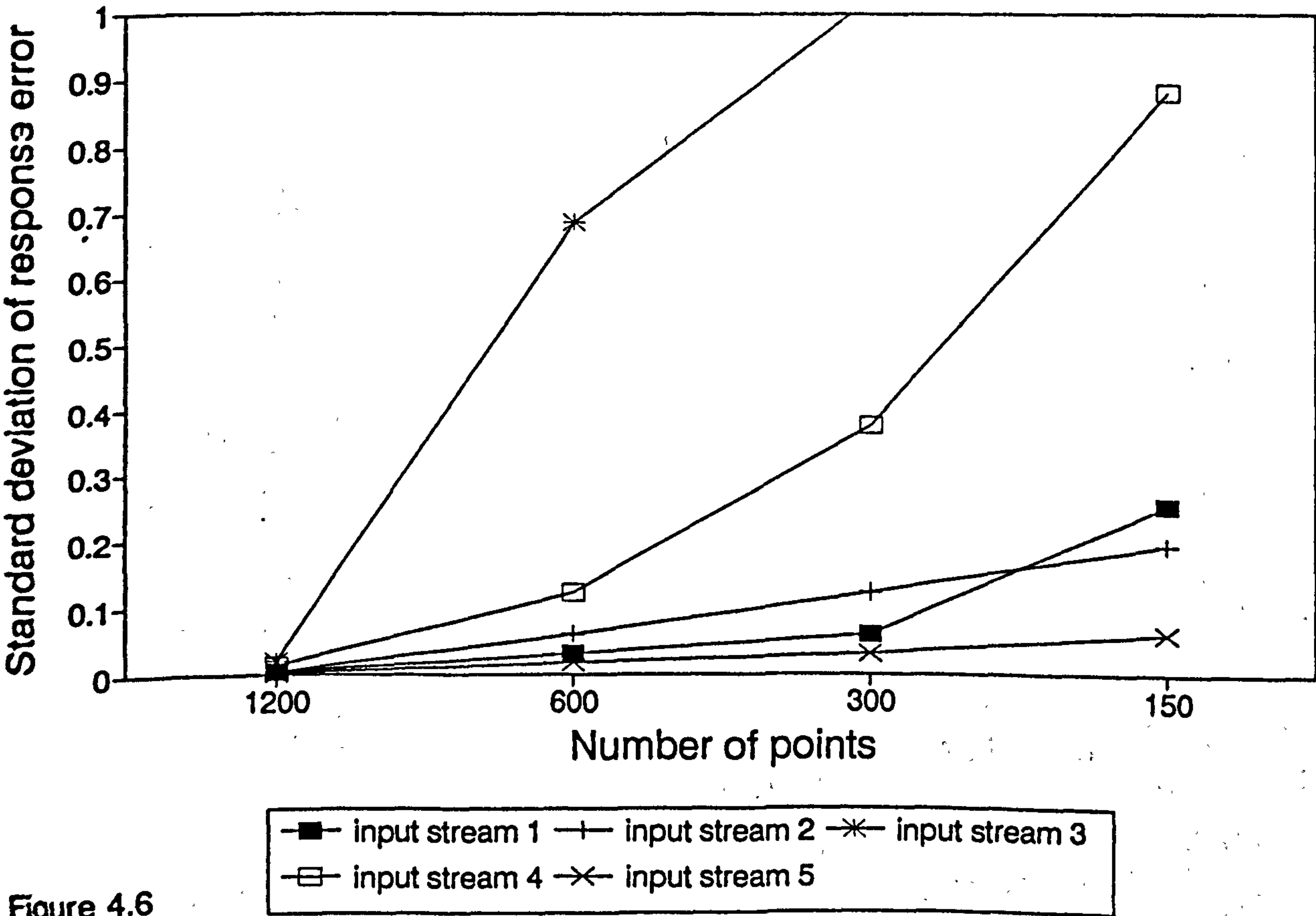


Figure 4.6

CHAPTER 5

TESTING THE METHOD

The work described in this chapter is aimed at determining whether a fitted error response model is of any value in determining the sources and causes of simulation error. A numerical experiment is devised to establish this, in which a series of test datasets are generated analytically. A simulation model is used to produce a set of building performance predictions which will be adopted as a quasi-truth model. The same model is then used to produce simulations which contain 'errors', caused by the perturbation of selected input parameters. The impulse response model is then used to characterise these errors in terms of the inputs driving the simulation, to see if the causes of the 'errors' can be correctly inferred. Before this test can be carried out a series of issues relating to the design of the numerical experiment have to be resolved. These include the choice of building to use for the test, which quantities are to be compared between the simulation runs, and which parameters should be varied to provide a comprehensive test of the method.

5.1 Analytical trials with SERI-RES

The building thermal model SERI-RES is a dynamic simulation code which allows the thermal performance of buildings to be predicted down to an hourly timescale. The model represents many of the phenomena contributing to the overall thermal performance of a building, but, in the interests of tractability, makes a number of simplifications. Specifically:

- the radiative and convective heat transfer processes within each zone of the building are combined. The temperature which the model calculates for each zone is therefore a combination of air and mean radiant temperatures. We will refer to this temperature simply as the zone temperature. It has been demonstrated [Haves, 1989] that under certain conditions the calculated zone temperature is a close approximation to the room air temperature,
- the model also combines convective and radiative heat flows on the outside of the building. Implicit in this approach is the fact that the model does not contain any representation of radiation to a cold sky: all external heat transfer is assumed to take place with the external dry bulb temperature, and
- the model does not adjust external heat transfer coefficients with wind velocity. The only

use of wind velocity is to adjust the building ventilation rate, an optional facility. This facility has not been used here, and the wind speed supplied in the meteorological data file plays no part in the simulation process.

For the purposes of this test, the performance of a small, single zone test room was simulated. The building described was actually one of a series at the EMC outdoor test facility [Martin and Watson, 1981] although the physical existence of the building is, of course, irrelevant for the purposes of the tests described here. However, real data from those buildings will be used to test the predictions of SERI-RES in the following chapter, making them an appropriate choice for a preliminary test of the analysis method.

Each test building at the EMC site consists of a semi-detached pair of test rooms. The party wall is well insulated, and can reasonably be modelled by assuming that there is no heat flow across its centre-plane [Lomas, 1988]. In this way the performance of single rooms can be simulated. Each building has an outer plywood skin, supported by a structural timber frame. The areas between the framing elements are filled with a glass fibre insulation quilt. The interior of the room is lined with plasterboard. The floor of each room consists of concrete slabs, supported on Styrofoam insulation. The rooms are tightly sealed, and can safely be assumed to have an infiltration rate of zero. The buildings themselves are supported clear of the ground, eliminating ground floor heat loss as a source of modelling uncertainty. The room windows are interchangeable, and for the test described here the room was assumed to be equipped with 1.5 m² of double glazing. Heating is provided by a small electric panel radiator. The rooms are described in more detail in Chapter 6.

SERI-RES provides the facility to tabulate a wide range of temperatures, radiation levels, heat flows, and energy balances for every part of a building. The next experimental design issue to be resolved is the choice of which of these quantities to use for the comparisons. Since the intention of this test is to show that the new method will ultimately be able to detect the causes of discrepancies between simulations and measured data it is vital to concentrate on quantities

which can be measured in a real empirical validation experiment. In practice these are generally air temperature, surface temperatures and surface heat fluxes, and these quantities will be considered here.

The model was used to predict the performance of a double glazed test room over a 50 day period. The model was supplied with real meteorological data, and with a randomised heater operation schedule. It was used to predict the test building zone temperature, and the temperatures and heat fluxes at the internal surfaces of the floor, walls and ceiling of the building.

The next stage in the experimental design process is to determine which parameters should be changed to generate the 'perturbed' simulation results. The thermal processes within the test room can broadly be broken down into conductive heat loss, the input of solar energy and its distribution around the rooms, and the consequent charging and discharging of the massive elements of the building structure. Of these processes, conductive heat loss can take place at a range of rates, from the high speed heat losses through, for example, glazing, to the slower fabric heat loss processes.

Clearly, to make the proposed test as comprehensive as possible it is desirable to test the diagnostic power of the method on all of these processes, over their entire range. To this end, five model input parameters were chosen for investigation. The first two described components of the conductive heat losses from the building. The first was the U-value of the building double glazing. The second was the thermal resistance of the layer of insulation used in the test room floor. The glazing heat loss is a relatively high speed process, but the floor insulation is beneath a massive floorslab. Hence it is to be expected that as well as the relative effects on the different quantities predicted, the dynamic effects of the two parameter changes will also be radically different. The next two parameters were related to solar processes. The first solar parameter varied was solar-loss, the amount of incoming solar radiation reflected back out of the test building as shortwave radiation. A change in this parameter affects the entire building

energy balance, and it would thus be expected to cause errors in all the quantities being predicted. The second solar radiation parameter chosen was the amount of radiation absorbed by the building floor. Radiation not absorbed in the floor is subsequently absorbed at the other building surfaces, and thus this parameter can be expected to have a less significant effect than solar-lost, as it does not affect the overall energy balance. Furthermore, its effect is expected to be concentrated at one surface rather than distributed throughout the building. The final parameter varied was a thermal mass parameter, the density of the room floorslab.

The next issue to be resolved was by how much the parameters should be varied. It is clear that as the variation in a single parameter is made smaller and smaller its effect will eventually become undetectable. However, in order to be useful, the proposed method must be sufficiently sensitive to detect the sort of parameter errors which are likely to be seen in real empirical validation experiments. To this end, the parameters described above were each perturbed by an amount roughly equal to the uncertainty which might typically exist in its measured or assumed value. This strategy has an interesting consequence in terms of the output uncertainty criteria for validity which were discussed in Section 2.4.2 of Chapter 2. The change in model output due to an error in a single parameter equal to the uncertainty in that error must inevitably be less than the total output uncertainty due to all the model parameters. Thus an output uncertainty analysis would fail to detect the parameter variations which have been used here, instead simply declaring the model 'valid'.

Table 5.1 summarises the parameters varied in the simulations, and gives the changes which were made to each one.

Process	Parameter	Nominal value	Perturbed value
Rapid heat loss	Glazing U-value	3.4 W/m ² K	3.8 W/m ² K
Low speed heat loss	Underfloor Styrofoam conductivity	0.027 W/mK	0.033 W/mK
Solar entering building	Solar-lost coefficient	0.20	0.24
Absorption of solar at room surfaces	Proportion of solar to floor	28%	38%
Passage of heat into and out of thermal mass	Floorslab density	2000 kg/m ³	1600 kg/m ³

Table 5.1: Parameter variations for analytical test of method

The differences between the results of the five simulations incorporating these perturbations and the base case simulation, in which all the parameters were set to their nominal values, were then generated. In all cases the difference was defined as the output from the perturbed simulation minus the output from the base case. Thus a positive difference implies that the perturbation increases the predicted quantity. These differences will be referred to as simulation 'errors', as this would be their rôle in an empirical validation exercise, even though in this context they do not actually represent faults in the simulation model. They are in general small. As an example, Figure 5.1 shows the base case prediction of the test room zone temperature, and the prediction when the Glazing U-value is altered by the amount given in Table 5.1. Figure 5.2 shows the corresponding change in floor surface temperature. Figures 5.3 and 5.4 show the corresponding results when the floor thermal capacity is varied. In all cases the changes are very small, and it is likely that a straightforward graphical comparison of the simulated and perturbed results would conclude that they were effectively the same, and that the perturbed model was 'valid'.

To confirm the relatively small effect of the parameter variations Table 5.2 shows the mean and standard deviations of the change in simulation model predictions when the first parameter, glazing U-value, is perturbed. In the table, as in the following discussion, the surface temperatures and heat fluxes from only the floor and a single, representative, wall are considered in the interests of brevity.

Quantity	Nominal mean value	Mean change	Standard deviation of change	Units
Zone temperature	26.3	-0.7	0.1	°C
Floor surface temperature	25.7	-0.7	0.1	°C
Floor surface heat flux	8.0	-0.3	0.2	W/m ²
Back wall surface temperature	25.7	-0.7	0.1	°C
Back wall surface heat flux	6.0	-0.3	0.1	W/m ²

Table 5.2: Sensitivity of predicted parameters to glazing U-value perturbation

5.2 Identification of sources of simulation error

For each parameter perturbation the discrepancies between the resulting predictions and those of the base case simulation were analysed using the computer program described in Section 4.4 of Chapter 4. The impulse response of these discrepancies to each of the four forces driving the test rooms was identified. Then the discrepancy was reconstituted by feeding the input streams through the identified responses, and the reconstructed sequence compared with the actual discrepancy observed. This serves to give a measure of how well the chosen linear time-invariant error model represents the discrepancy. Finally, the relative contributions to the reconstructed error from each of the driving forces were isolated, and their relative contributions to the overall discrepancy determined.

Before attempting the interpretation of these results it is instructive to consider the expected shape of the error impulse responses for different types of error.

The speed at which an error mechanism operates should be immediately apparent from its impulse response: this is, after all, plotted in the time delay domain. Errors which manifest themselves rapidly will be characterised by responses close to the origin, lower speed errors will produce responses which peak at higher time delays.

To determine the form of the error impulse response associated with different kinds of modelling error consider first an error which increases the heat loss of a building. Recall that the error impulse response to each model input charts the way in which the resulting error evolves when an impulse excitation is applied to each model input. In the case of an impulse in heating input to the building the resulting temperature rise will at all times be reduced. The response to an impulse in solar radiation will be similar, with the resulting temperature rise again being reduced by the modelling error. Finally, if external temperature undergoes an impulse change the error response of the temperatures within the building will respond with an increase: a closer coupling being generated by the increased building heat loss to ambient. It is clear that a modelling error which decreases building heat loss will give rise to impulse responses with the characteristics listed above, but with signs reversed. In summary, errors in building heat loss will be expected to lead to an error structure which displays contributions from heater operation, solar radiation and external ambient temperature. The signs of these responses should allow the sign of the error (whether it acts to increase or decrease building heat loss) to be determined.

Consider now the effect of an error in the modelling of the thermal storage within the fabric of a building. In the situation where the error causes the mass of the building to be underestimated the response of the building temperature error will initially be positive: the reduced mass causes the temperature to rise more quickly than expected. Subsequently, however, the reduced energy placed into storage causes the building to be cooler than expected. Errors in the modelling of

thermal mass will therefore be characterised by responses which change sign. The point in time at which this change occurs will be determined by the mass of the element causing the error, the process being slower for high mass elements.

The above discussion has been couched in terms of global building parameters, heat loss and thermal mass, and a global building temperature. In practice it will be of interest to determine the particular building element which is producing the simulation error. The most direct route to this is via mechanism level data, for example the temperature of or heat flux into a given building element. It is intuitively obvious that the measurement which shows the largest response is likely to be the one closest to the source of the error. For example, if floor temperature and heat flux measurements display the largest error response the source of the error, be it heat loss or thermal capacity, is likely to be in the modelling of the construction of the floor.

In the light of these very general observations, the results of applying the error response analysis to the perturbed simulations defined above will now be examined.

5.2.1 Conductive heat loss effects

Consider first the simulation run carried out with the room glazing U-value perturbed. Figure 5.5 shows the response of the discrepancies in predicted air temperature to the driving forces. Figures 5.6 and 5.7 show the responses of floor surface temperature and heat flux errors. Figures 5.8 and 5.9 show the corresponding responses for the back wall. In all cases the response of the errors to windspeed is zero, as expected from the description of SERI-RES in Section 5.1.

The impulse responses shown on Figures 5.5 to 5.9 would normally have different units. For example the response of a temperature error to external temperature would have units $^{\circ}\text{C}/^{\circ}\text{C}$, those to solar radiation $^{\circ}\text{C}/(\text{W}/\text{m}^2)$, to windspeed $^{\circ}\text{C}/(\text{m}/\text{s})$ and to heater operation $^{\circ}\text{C}/\text{W}$. However, the numerical magnitude of solar radiation in W/m^2 is very much larger than external

temperature in °C, and as a result the impulse response will generally be very much smaller, making comparison difficult. The approach which has been adopted on these and subsequent figures is to scale each response by dividing by the mean value of the associated input. This facilitates comparisons between responses in terms of their overall effect on the total prediction error.

Next, Figure 5.10 shows the reconstituted error in zone temperature, obtained by convolving the each of the driving forces with the estimated responses of Figure 5.5. Figure 5.11 shows the result of combining the reconstituted errors due to each source, and also shows the actual error. The figure demonstrates that the proposed linear time-invariant error model provides a good reconstruction of the observed 'error' sequence. The main source of discrepancy is a small offset, which corresponds to the constant c introduced in Section 4.3 of Chapter 4. Figure 5.12 shows the relative contributions from each of the driving forces, and the section of the error sequence which is unaccounted for. The figure confirms that the assumed linear time-invariant model accounts for the majority of the observed error in all the quantities predicted, the variance of the part of the error not accounted for being indiscernibly small.

Reference to Figure 5.5 shows that the response of the error to external temperature is by far the largest of the responses generated. However, this conclusion is apparently contradicted on Figure 5.12 where, although the contribution from external temperature is significant, it is smaller than those from the room heat source or solar radiation. The reason for this was identified in Section 4.5 of Chapter 4 and lies with the slowly changing nature of this driving force. Although the impulse response associated with it is large, it has a significant oscillatory component. Because external temperature changes only slowly, the effect of convolving it with this oscillatory response is one of smoothing the response, reducing the magnitude of its contribution to a lower level than might initially have been expected. Clearly, it is important that the impulse responses are not interpreted in isolation - their significance may depend strongly on the character of the driving force which underlies them.

The systematic interpretation of these results begins by considering which driving forces affect the error, then examines the effect that the perturbation has had on overall building heat loss, looks at establishing the location of the perturbation within the room, and finally moves on to consider the dynamic characteristics of the error:

- Figure 5.12 reveals that the errors in all the quantities predicted show contributions from the room heater operation, external temperature and solar radiation. This indicates that the source of the discrepancies cannot be attributed directly to any one of these mechanisms: it must be linked to all of them. From the general discussion of the previous section we conclude that the error is likely to be in building heat loss.
- The response of zone temperature to heater operation shown on Figure 5.5 reveals that the change associated with the perturbation is negative, implying lower zone temperatures and hence that the error has increased building heat loss.
- Examining the impulse responses of the surface temperatures to heater power on Figures 5.6 and 5.8 reveals that these are negative, implying that the error becomes more negative as the heater is operated, an observation which is consistent with the above observation that the perturbation has increased building heat loss.
- When the heat flux error responses to heater power are examined, on Figures 5.7 and 5.9, these are also found to be negative, implying lower heat fluxes out through surfaces when the heater is operated in the perturbed model. Equivalently, the change in the surface heat fluxes when the perturbation is effected is in an inward direction. Figure 5.13 shows this diagrammatically. Since the building heat loss has actually increased, the increase must be through a building element other than these surfaces. Generally the analysis would have considered heat fluxes at all opaque surfaces, although here only two are considered for the sake of brevity. The only remaining candidates for the increased building heat loss observed are therefore the window, and ventilation heat loss.
- Finally, the impulse responses shown on Figures 5.4 to 5.9 indicate that the effect occurs over a short timescale. The error response of zone temperature to heater operation, shown on Figure 5.5, peaks after only 2 hours. The effect is even more pronounced in the corresponding responses of the surface heat flux errors, which peak after only 1 hour for

the floor on Figure 5.7, and instantaneously for the back wall on Figure 5.9. The surface temperature errors peak after 3 to 6 hours, as the mass of the surface itself introduces additional delay into these responses. It is therefore clear that the extra heat loss of the building is occurring via a relatively high speed heat flow path. Again there are in general two such mechanisms in a building: glazing heat loss and ventilation.

Since the building modelled here is unventilated we conclude from the above discussion that the source of the observed discrepancies must be an increase in glazing heat loss. If the building was ventilated the above discussion would not have been able to separate the possibility of an error in the specified ventilation rate from an error in glazing heat loss. The reason for this is simple: the two heat flow paths are in parallel, and they have similar dynamic characteristics. One way of isolating them would be to inspect, for example, the surface temperature of the glass, which would provide mechanism level information on whether the glazing heat flow was the quantity in error. Alternatively, and this is the approach which has implicitly been adopted here, the problem can be eliminated at the experimental design stage by eliminating one or other of the mechanisms from the test.

The second heat loss parameter was the conductivity of the Styrofoam used as insulation beneath the concrete floor of the room. Figures 5.14 to 5.18 show the impulse responses of the errors in predicted zone temperature, and floor and back wall surface temperatures and heat fluxes. Figure 5.19 shows the relative contributions to the error responses for each of the seven quantities predicted. As before, each of the errors shows components due to heater operation, external temperature and solar radiation. However, unlike the previous such figure, it is clear from Figure 5.19 that the source of the error is concentrated in the floor of the room. This is apparent from both the temperature and heat flux results.

Reference back to the impulse responses shows a different pattern to that observed when the glazing heat loss was perturbed. The response of zone temperature error to heater operation shown on Figure 5.14 is again negative, as is the response of the floor temperature on Figure

5.15. However, the floor heat flux response shown on Figure 5.16 is now clearly positive, implying that the effect of the perturbation is to increase the heat flux into the floor surface when the heater is operated. On the back wall the response of surface temperature error is again negative, implying that the perturbed room is cooler than the base case when the heater is operated. However, the heat flux response is now also negative, implying that the perturbation decreases the flux into the surface of the back wall. Taken together these responses clearly indicate that the perturbation has the effect of increasing the heat loss through the floor, whilst at the same time causing an inward heat flow at the back wall, which is typically ten times smaller than the change observed at the floor. Again, these changes in internal surface heat fluxes are shown diagrammatically on Figure 5.20, and this time they point clearly to an increase in floor heat loss. The inward change in flux observed at the back wall is a real consequence of this. The increased floor heat flux reduces the floor surface temperature, and this in turn increases the convective and radiative heat flux from the back wall to the floor.

The impulse responses further reveal that the effect of the perturbation is delayed in time, considerably more than in the previous example. This is again particularly clear in the case of the responses to auxiliary energy input. The floor heat flux error is seen to reach a peak after approximately 6 hours, floor temperature error peaks after 8 hours, and zone temperature error after 12 hours. This clearly demonstrates that the source of the errors must be embedded beneath the thermal mass of the floor.

5.2.2 Solar radiation effects

The first solar radiation parameter explored was solar-lost. Figures 5.21 to 5.25 show the impulse responses of the errors in predicted zone temperature, and floor and back wall surface temperatures and heat fluxes. Figure 5.26 shows the relative contributions to the overall error in each quantity. The figure differs from those examined previously, in that it indicates that all the prediction errors are quite clearly due to a single source, solar radiation.

Having established that solar radiation is the source of the discrepancies, reference back to the

impulse responses gives an indication of the sign and dynamics of the effect. It is clear from the temperature prediction impulse response curves that the response of the prediction error to solar radiation is in all cases predominantly negative: implying that the prediction error becomes more negative as solar radiation levels increase, or that the perturbed simulation result falls further below the base case as solar radiation increases. All the responses are seen to be relatively high speed, implying that the effect must be occurring inside the room, rather than occurring on the outside and being communicated to the room by conduction through the fabric, since the latter process would introduce significant time delays.

The conclusion is thus that the overall amount of solar radiation reaching the room interior is being reduced. This must be due either to a reduction in the transmission of the glazing system, or an increase in the amount of solar radiation being reflected back out of the room.

Turning now to the second solar parameter perturbed, solar distribution, Figures 5.27 to 5.31 again show the error responses, and Figure 5.32 shows the relative contributions to the reconstituted error sequences. As for the previous case, Figure 5.32 demonstrates conclusively that the source of the errors is a solar radiation parameter. However, careful comparison of the relative magnitudes of the errors in the floor, back wall and ceiling surface temperatures reveals that the errors are this time concentrated at the floor surface.

Examining the relevant impulse responses reveals that the floor temperature and heat flux error responses to solar radiation are positive. Following the reasoning used above, this implies that more solar radiation is reaching the floor in the perturbed simulation. However, the responses of the back wall temperature and flux reveal that they are of the opposite sign. Thus more radiation is reaching the floor, but less is reaching the other surfaces of the room. This is in contrast to the previous case, where all the impulse responses indicated a reduction in the solar radiation input, and clearly points towards an error in the distribution of solar radiation within the room.

5.2.3 Thermal capacity effects

The effect of variations in building thermal capacity was explored by changing the density of the concrete floor. Figures 5.33 to 5.37 show the impulse responses of the errors in zone temperature and floor and wall surface temperatures and heat fluxes. Figure 5.38 shows the relative contributions to each error from each input. It is clear that the major errors occur in the predictions of the quantities associated with the floor, and that there are contributions to that error from heater operation, external temperature and solar radiation. Examining the associated temperature impulse responses shows that in all cases the error response is first positive, and becomes negative after 10 to 12 hours. Thus the response to an impulse input of energy is first an overprediction of room temperature, followed by subsequent underprediction. As described above this is in keeping with a misrepresentation of the building thermal mass. When the responses of the surface heat fluxes are examined the floor responses are found to indicate initial underprediction, followed by overprediction. This is again consistent with a reduced floor mass - the reduced mass rises in temperature more quickly, and thus the heat flux into it is reduced. The surface heat flux at the back wall displays a much smaller effect, the flux being initially overpredicted, and subsequently underpredicted. This second order effect can be explained in terms of the effect on zone temperature already analysed: the initial overestimate of zone temperature causing a higher heat flux into the back wall.

In conclusion, these responses point to an error due to the thermal processes within the floor of the room, and indicate that the process causing the change is likely to be related to a thermal capacity effect.

5.3 Correlation and causality

It has been demonstrated that the sources of discrepancies between two simulation runs can be characterised and, in broad terms, identified. It thus appears that we have a powerful tool with which to analyse discrepancies between the predictions of a simulation model and measurements of the corresponding reality.

An issue which remains is whether the ability to explain errors in terms of the chosen error model necessarily implies that the actual source of the errors has been correctly identified. This is equivalent to determining whether a correlation between a simulation error and a driving force actually indicates a causal connection.

There are at least two situations in which a method of the type described could lead to erroneous conclusions about the source of an error. Not surprisingly, they correspond to violations of the linearity and time invariance assumptions respectively:

- a non-linear relationship may exist between simulation error and a driving force. For example, there may be an error in predicted zone temperature which is proportional to windspeed squared. The analysis employed here cannot represent this effect, and will attempt to parameterise it in terms of the linear model postulated. If windspeed squared happens to be more closely correlated to, say, solar radiation rather than to windspeed itself, then the source of the error will be incorrectly indicated.
- a time varying effect may cause an error to be present at some times of day, but not at others. Since some of the forces driving the test building have a strong diurnal component, this may again give rise to incorrect indications.

Such rather pathological sources of error may seem unlikely in the context of a carefully controlled empirical validation experiment. However, it must be remembered that the discrepancy between measured and simulated performance contains contributions both from the measured performance of a real building, but also from the simulation model. Whilst the processes within the building are at least constrained to follow physical laws, the predictions of a simulation model are effectively unconstrained, in that very minor errors in coding have the ability to produce the most unlikely of errors.

5.4 Conclusions and guidelines for the design of validation experiments

The analysis described in this chapter has demonstrated that an identified error response model can be used to diagnose the reasons for discrepancies between two sets of building performance

data. The technique has been shown to be sufficiently sensitive to analyse discrepancies of the size which are likely to be observed in an empirical validation experiment. Based on the physics which underlies the heat transfer processes in a simple building a series of guidelines have been developed which allow the nature of a simulation error (or in this case, model input parameter variation) to be deduced from the identified impulse response model. An error can be characterised as coming from a heat loss path or from the incorrect representation of a thermal mass process. The dynamic nature of the responses can reveal whether the error is associated with a fast thermal process, or whether it is due to the representation of an element embedded in the mass of the building. In order to identify the physical location of an error source it is also useful to have measurements of 'mechanism level' quantities, generally the surface temperatures and heat fluxes of individual elements.

In fact, a number of conclusions about the design of model validation trials and the subsequent identification and interpretation of error impulse responses can be drawn from the experiences of this chapter. When designing an experiment:

- surface heat flux data is generally a more sensitive indicator of the source of simulation errors than the corresponding surface temperature, and thus measuring heat fluxes will yield a more useful data set, and
- some processes (for example glazing heat loss and ventilation) may be very difficult to separate. Wherever possible, one of the processes should be 'designed out' of a validation experiment. For example, in an experiment designed to assess the quality of glazing heat loss modelling, ventilation should either be extremely accurately measured and modelled, or eliminated.

When interpreting the resulting error impulse responses:

- oscillation at the beginning of a response may indicate that the variable driving that response does not contain sufficient high frequency variation to allow the response at short time delays to be identified, and
- it is important that the interpretation process first considers the relative contributions to

reconstituted error as well as at the impulse responses themselves. An apparently large impulse response does not necessarily imply a significant contribution to the error. Once a contribution has been established, however, examining the appropriate impulse response can yield useful information about the error process dynamics.

5.5 The rôle of the new technique in the model testing process

The new tool which has been developed and tested provides a method for establishing the contributions of individual model inputs to discrepancies between model predictions and a corresponding reference dataset. In principle these data streams could be from any pair of sources, for example in a verification test they might be the output of the model under test and a 'perfect' implementation of selected algorithms from that model. In practice the technique is unlikely ever to be used in this context, for reasons which will now be outlined.

As discussed in Chapter 2 implementation, or verification, tests will never involve empirical data. In this situation a wide range of straightforward techniques can be employed to facilitate the comparisons between the model under test and the reference implementation. For example:

- building components or materials can be given idealised properties to remove selected aspects of their representation from the test. For example, a wall might be given a surface emissivity of zero in order to eliminate radiative exchange whilst the representation of convective processes is being tested, or
- selected inputs to the model can be artificially set to zero to remove their influence to determine the correct processing of remaining inputs. For example solar radiation might be 'switched off' whilst the representation of heat loss by conduction is tested.

Additional tools which allow the contributions of individual model inputs to errors to be deduced are therefore unlikely to be required in verification tests.

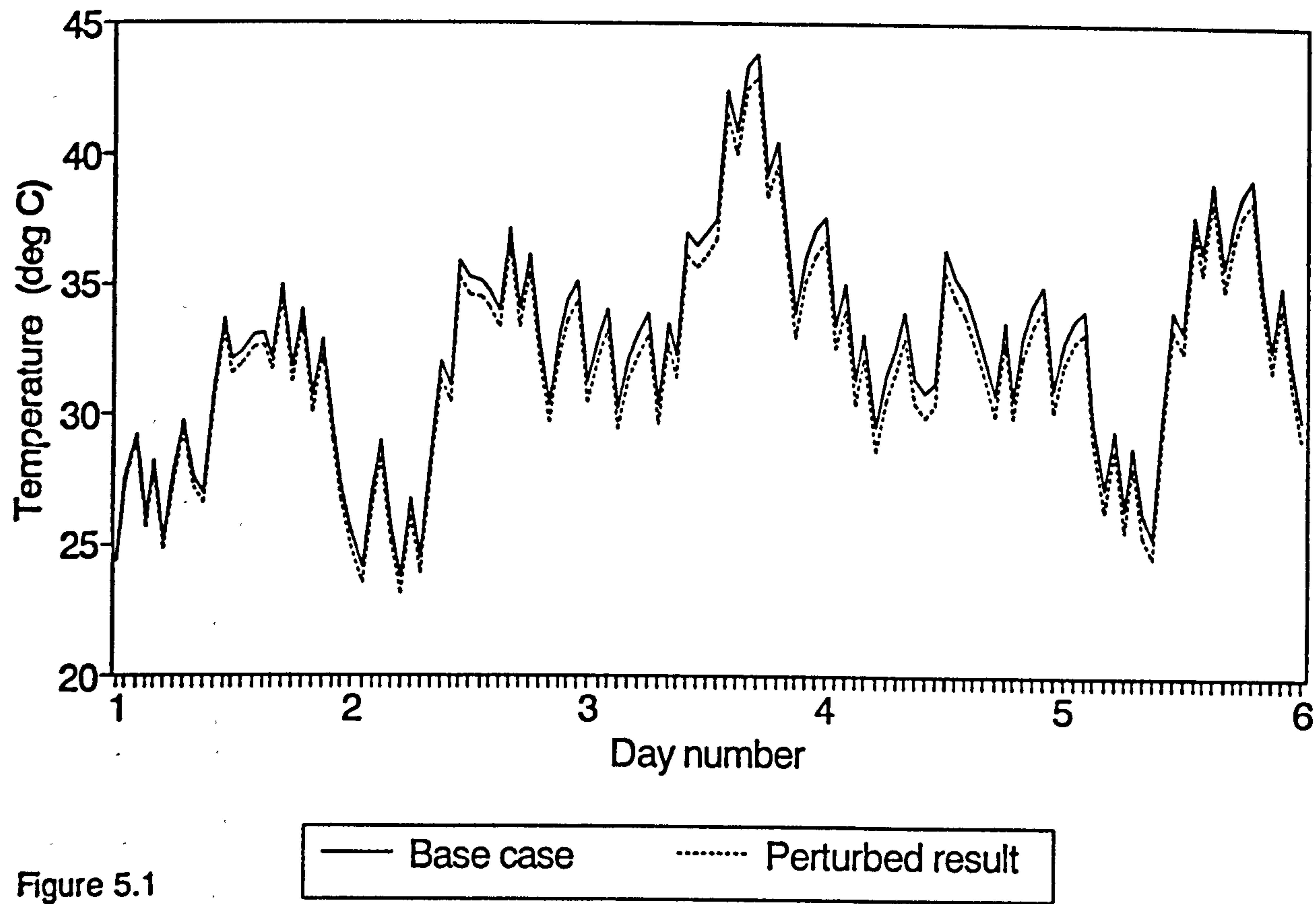
By contrast, tests of algorithms or empirical validation experiments aimed at whole model testing will inevitably require empirical data. In this situation the facility to isolate specific

processes will often not be available. Good experimental design will require that wherever possible unwanted effects should be minimised to emphasise the process under investigation. However physical constraints will often prevent these effects being completely eliminated in the way that they can be within an analytical test. For example, real surfaces can be given low emissivities to minimise radiation exchange, but materials with zero emissivity do not exist. Likewise, it may not be possible to eliminate solar radiation from an experiment, if the longwave exchanges between the external surfaces of a building and the sky are to be retained. Thus the nature of physical experiments almost always means that discrepancies between the algorithm under test or the whole model being validated contain contributions from a number of sources. It is in analysing the results from these tests that the new technique developed here should find general application.

FIGURES

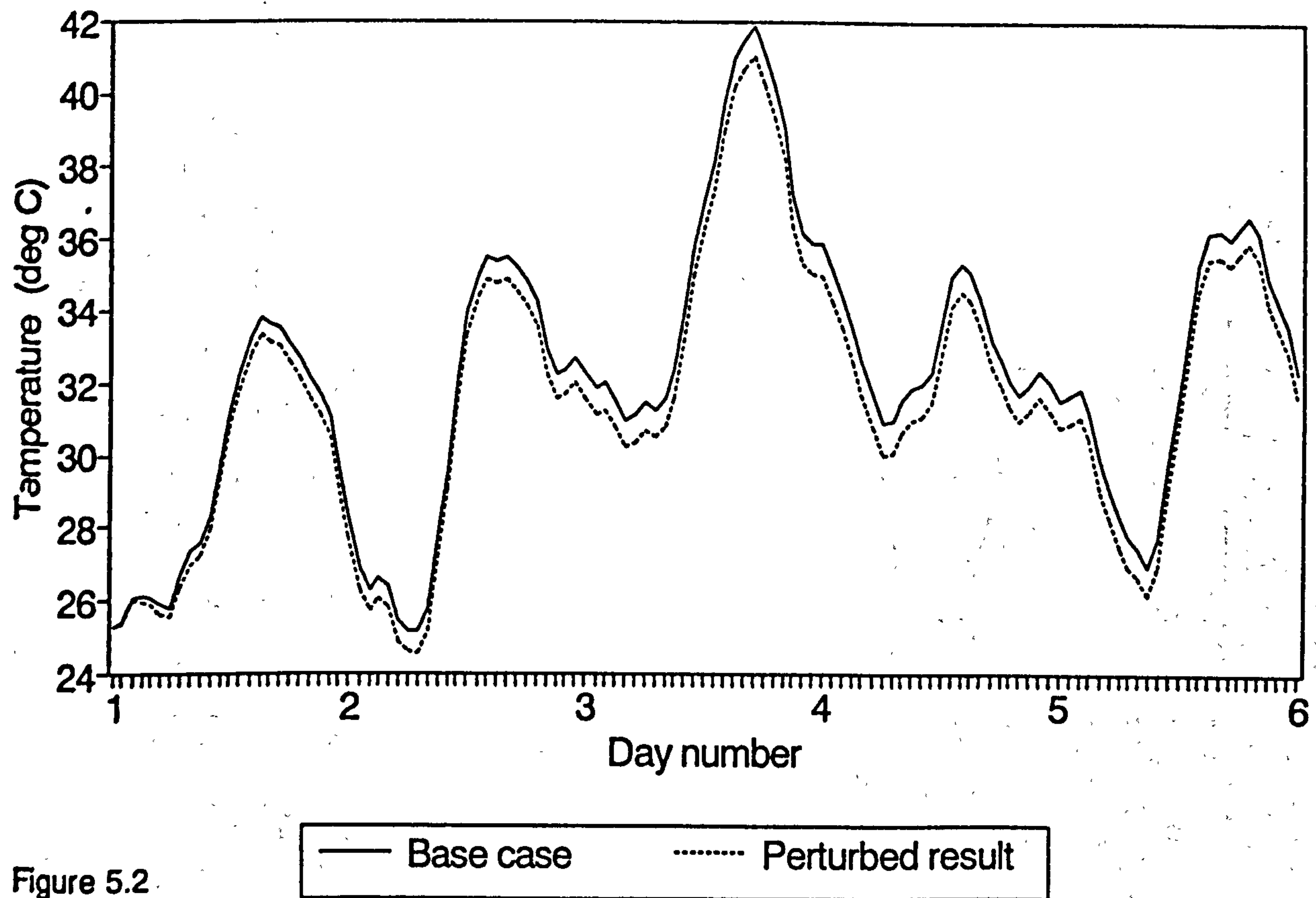
Effect of perturbing test room Glazing U-value

Zone temperature prediction



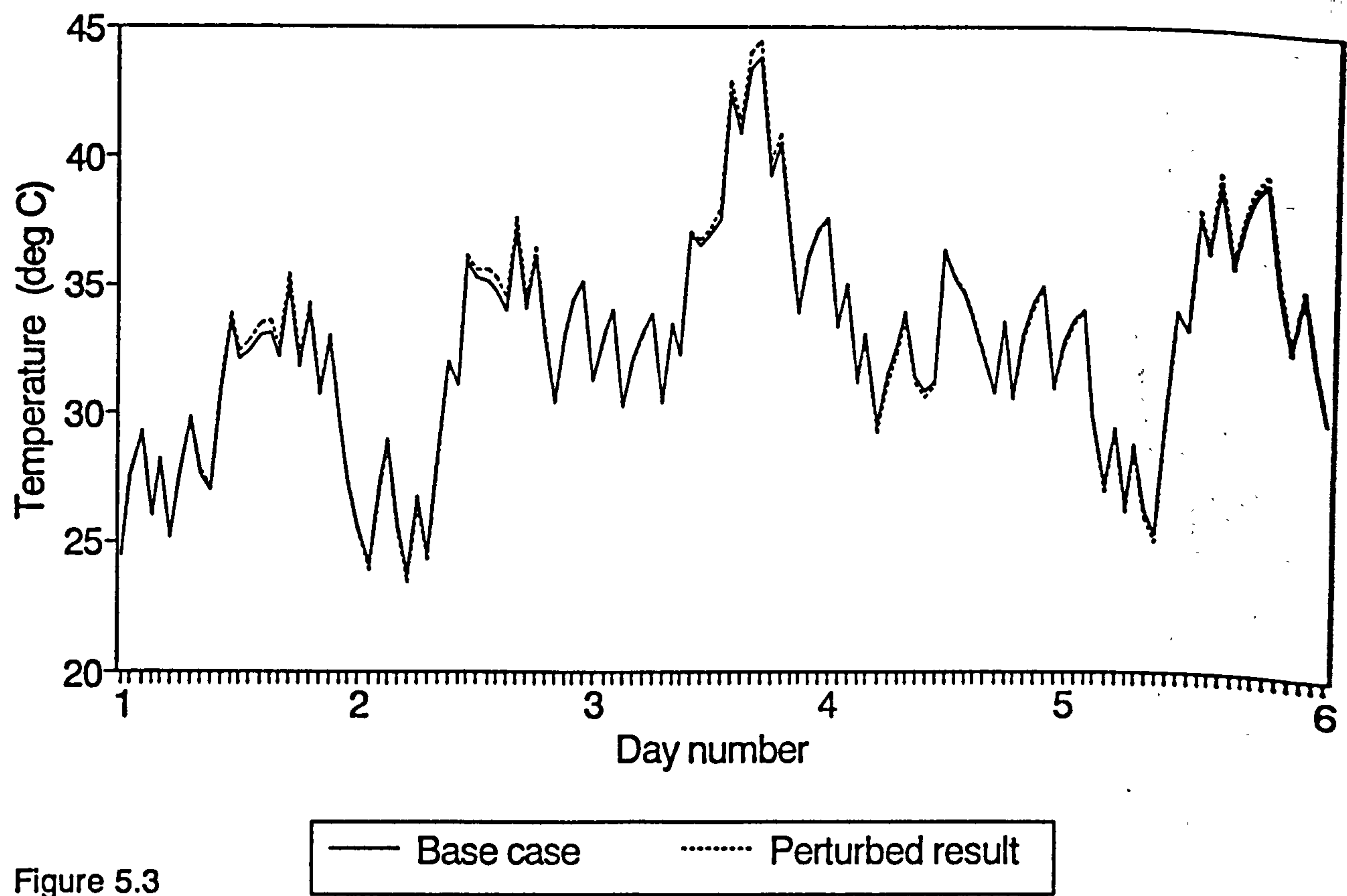
Effect of perturbing test room Glazing U-value

Floor surface temperature prediction



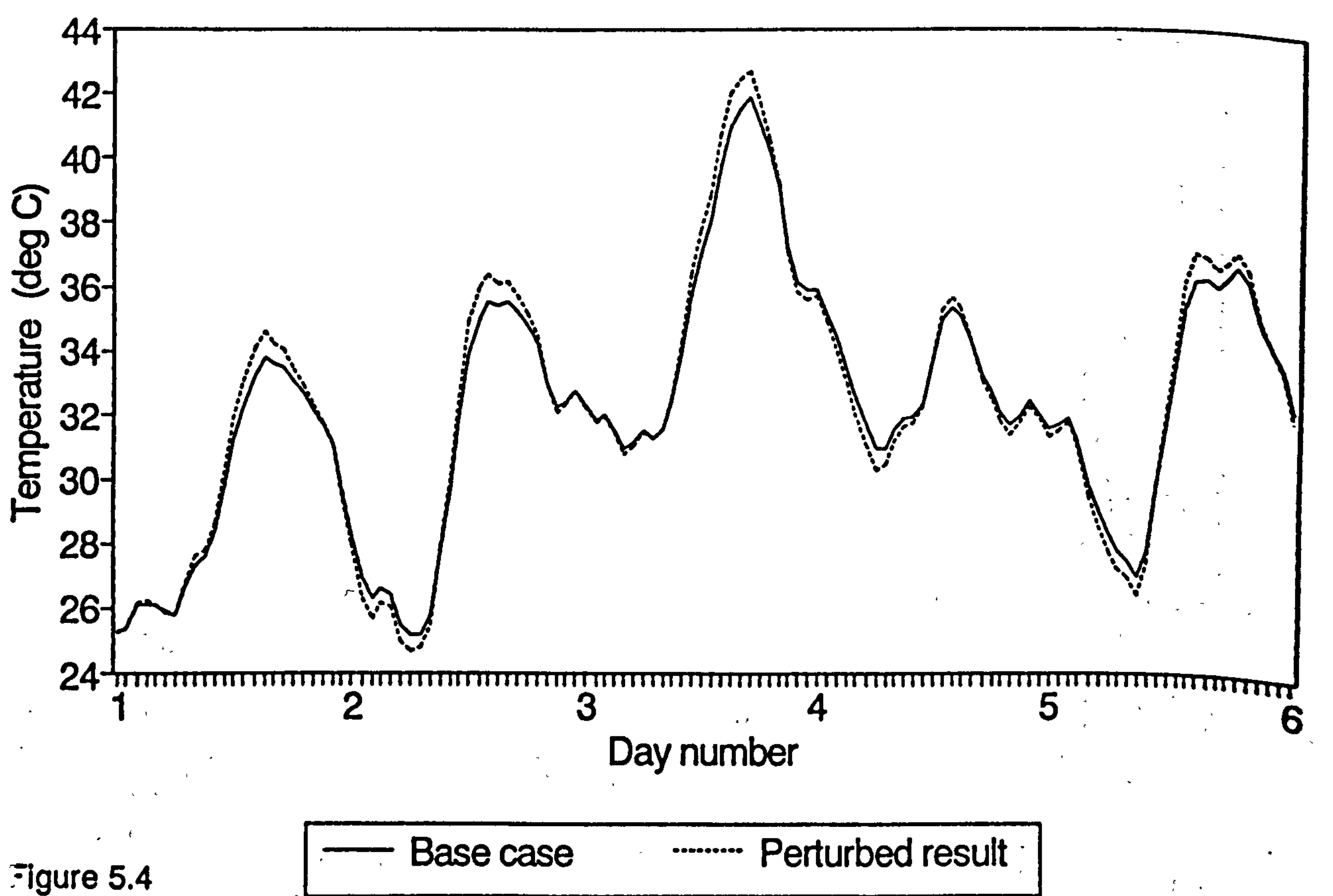
Effect of perturbing test room Floorslab density

Zone temperature prediction



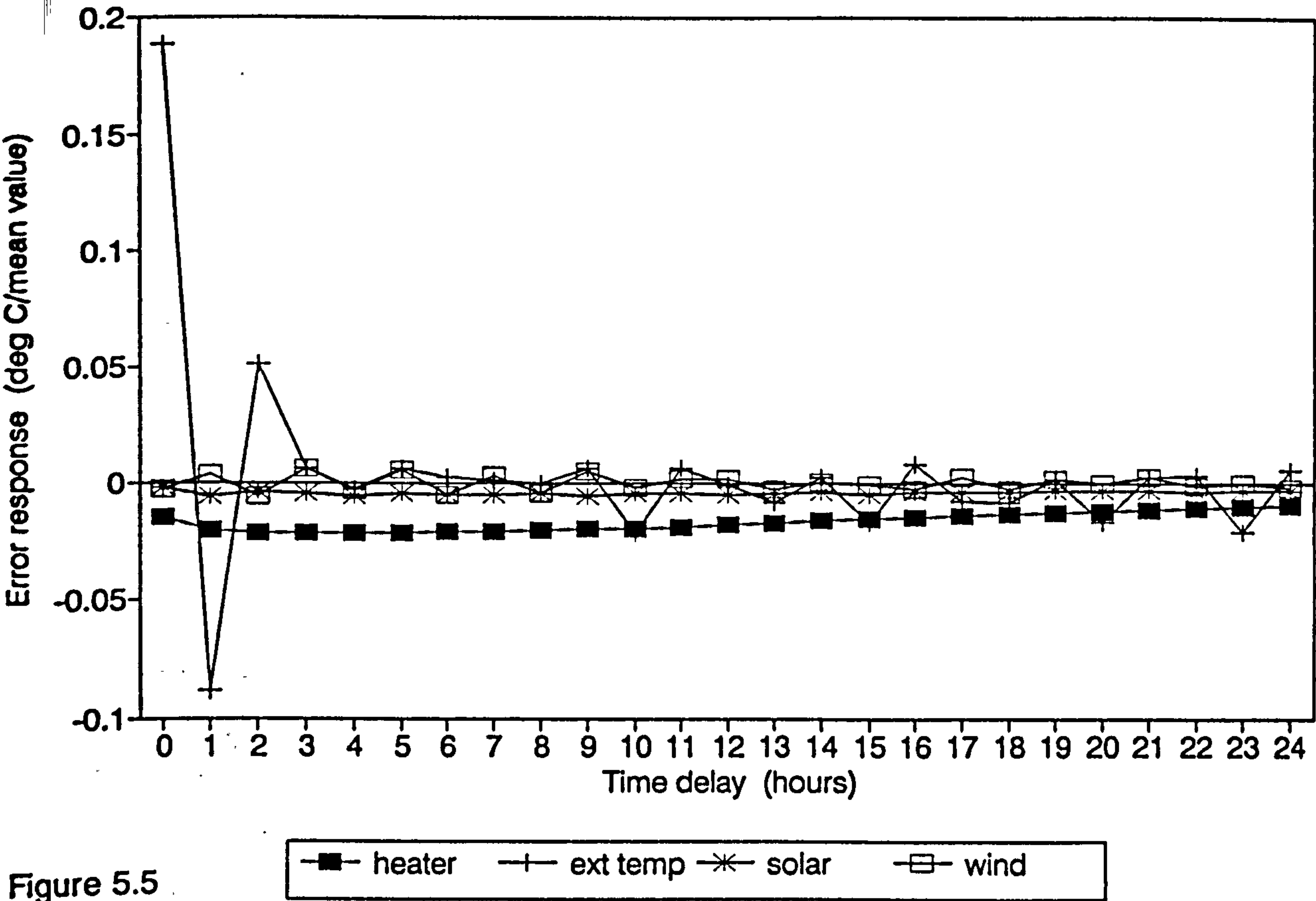
Effect of perturbing test room Floorslab density

Floor surface temperature prediction



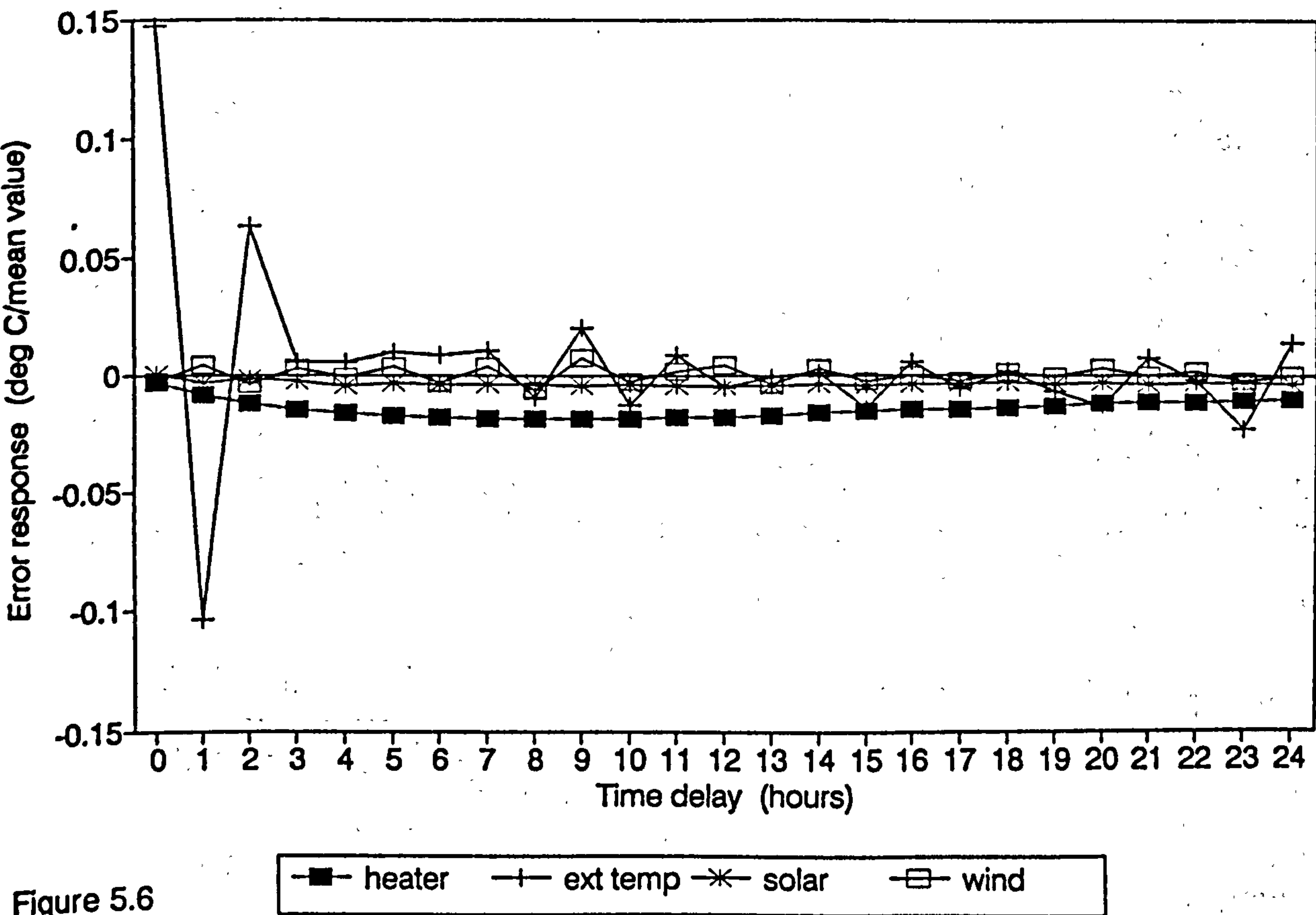
Sensitivity to error in Glazing U-value

Zone temperature prediction error responses



Sensitivity to error in Glazing U-value

Floor temperature prediction error responses



Sensitivity to error in Glazing U-value

Floor heat flux prediction error responses

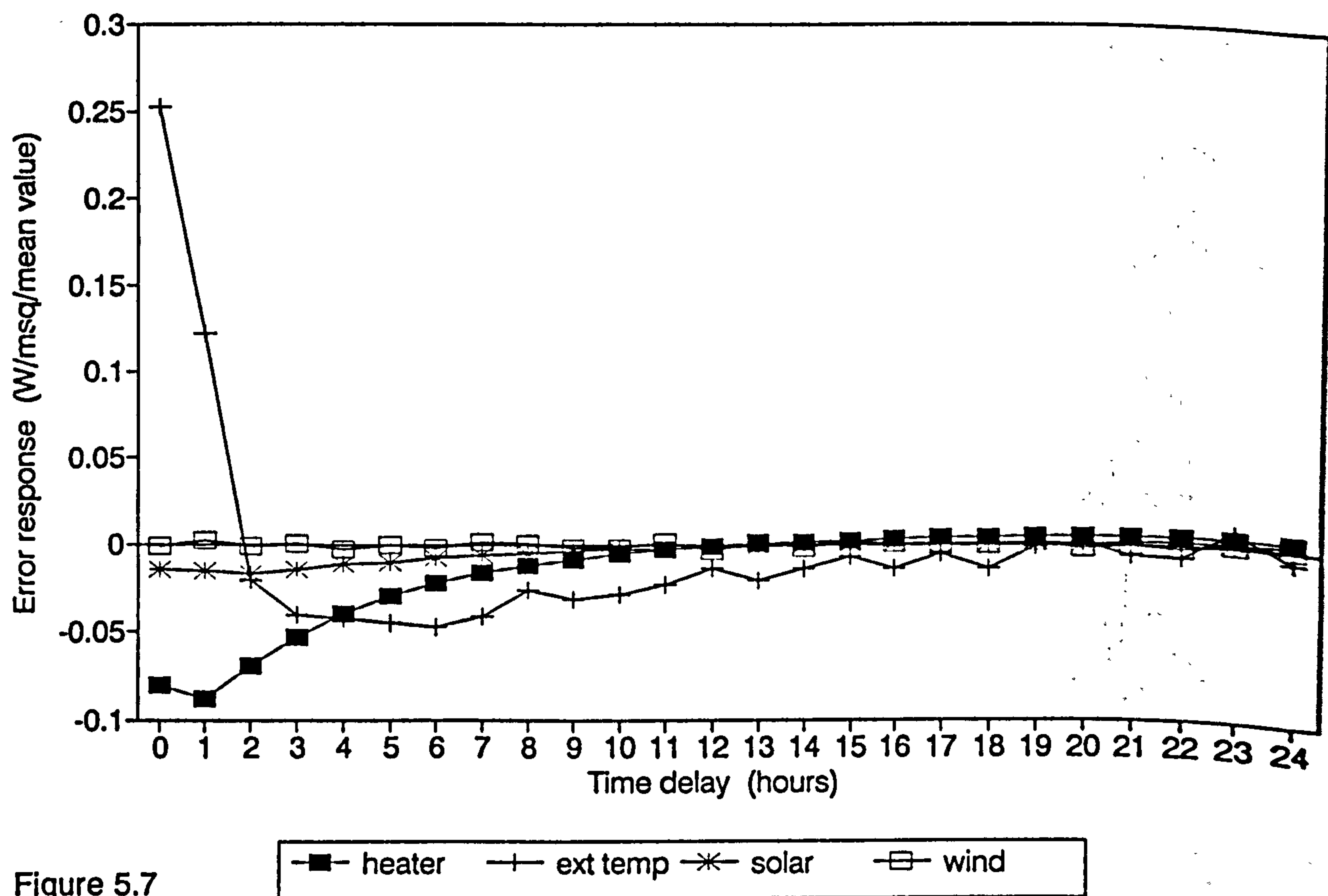


Figure 5.7

Sensitivity to error in Glazing U-value

Back wall temperature prediction error responses

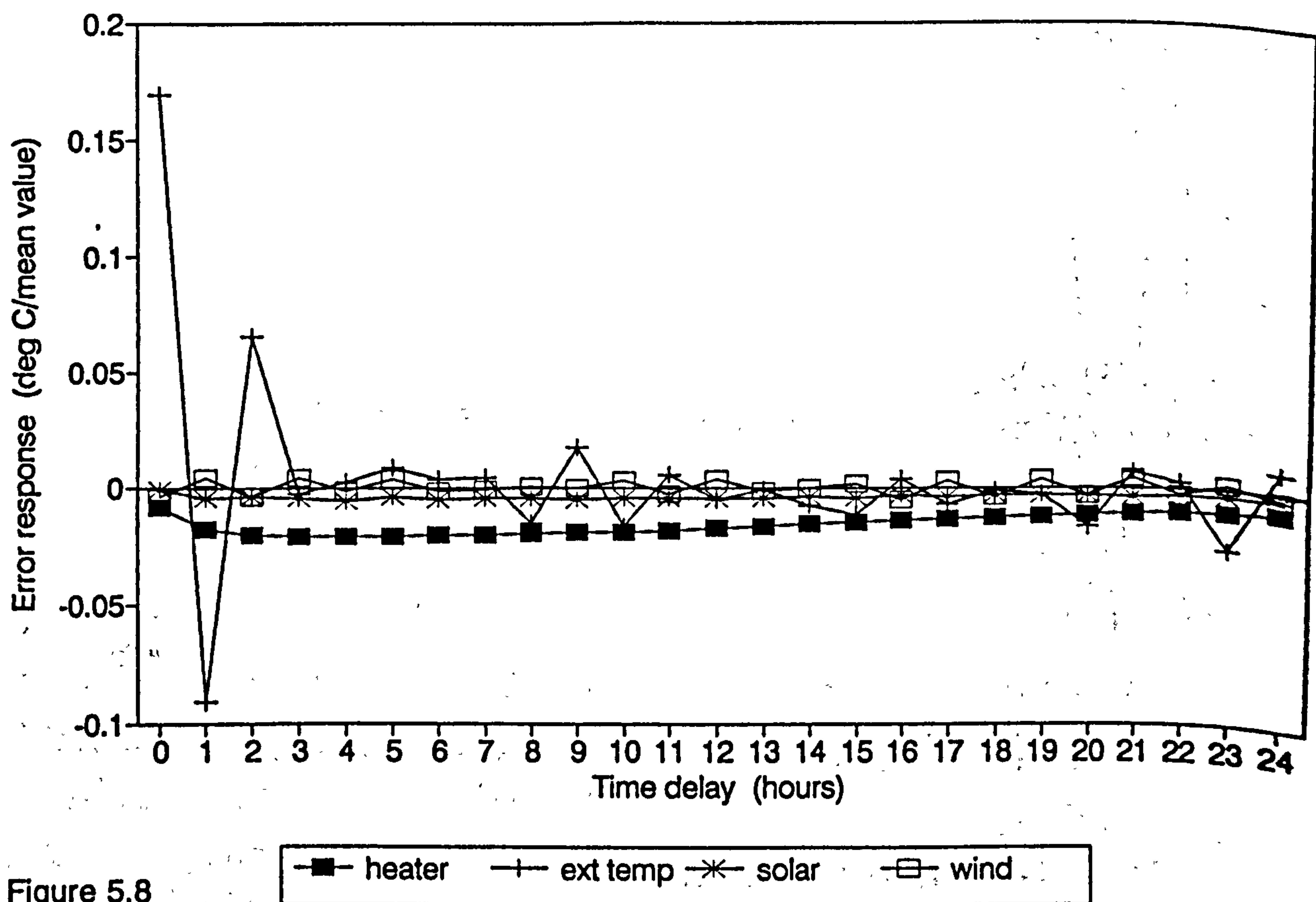
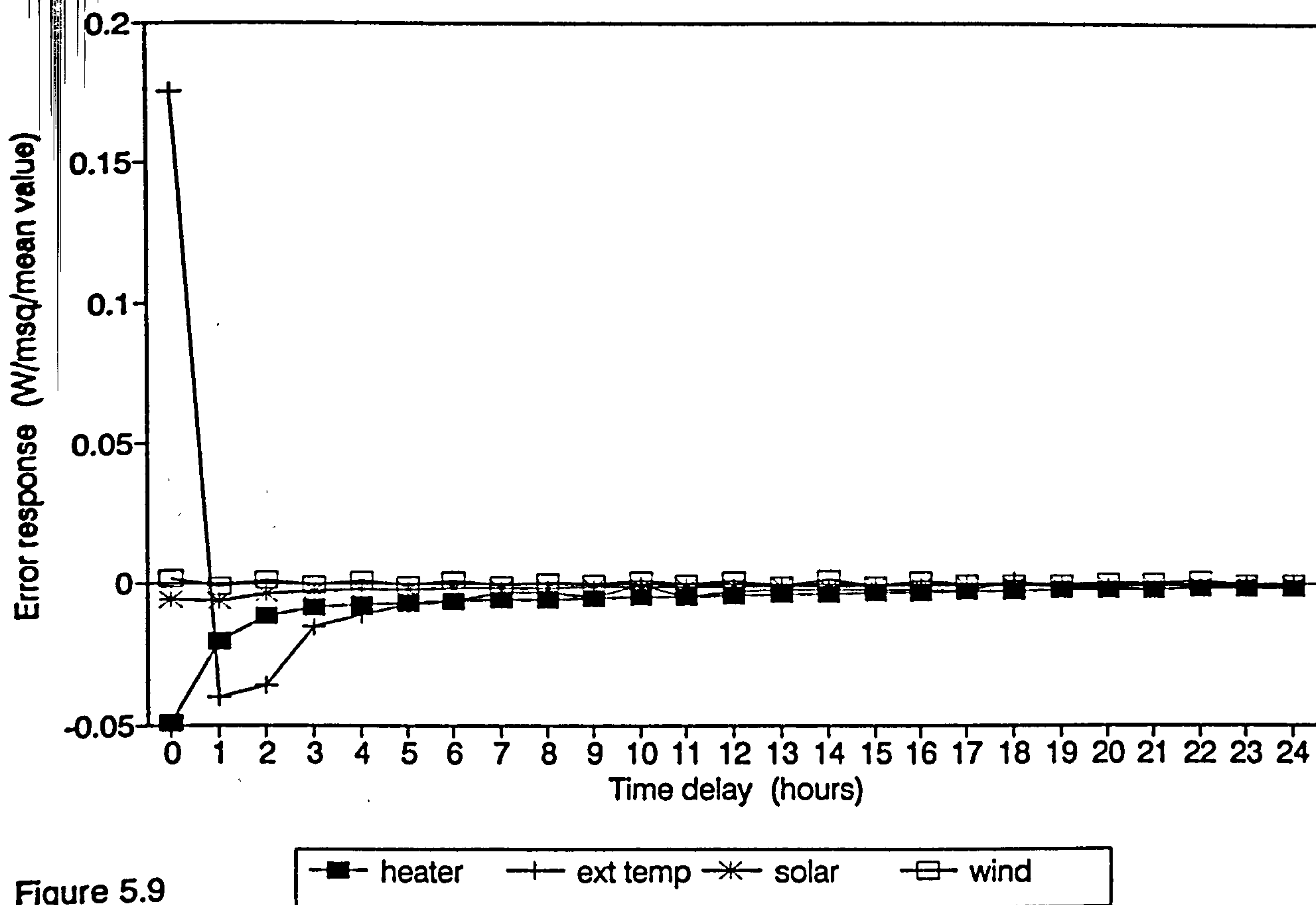


Figure 5.8

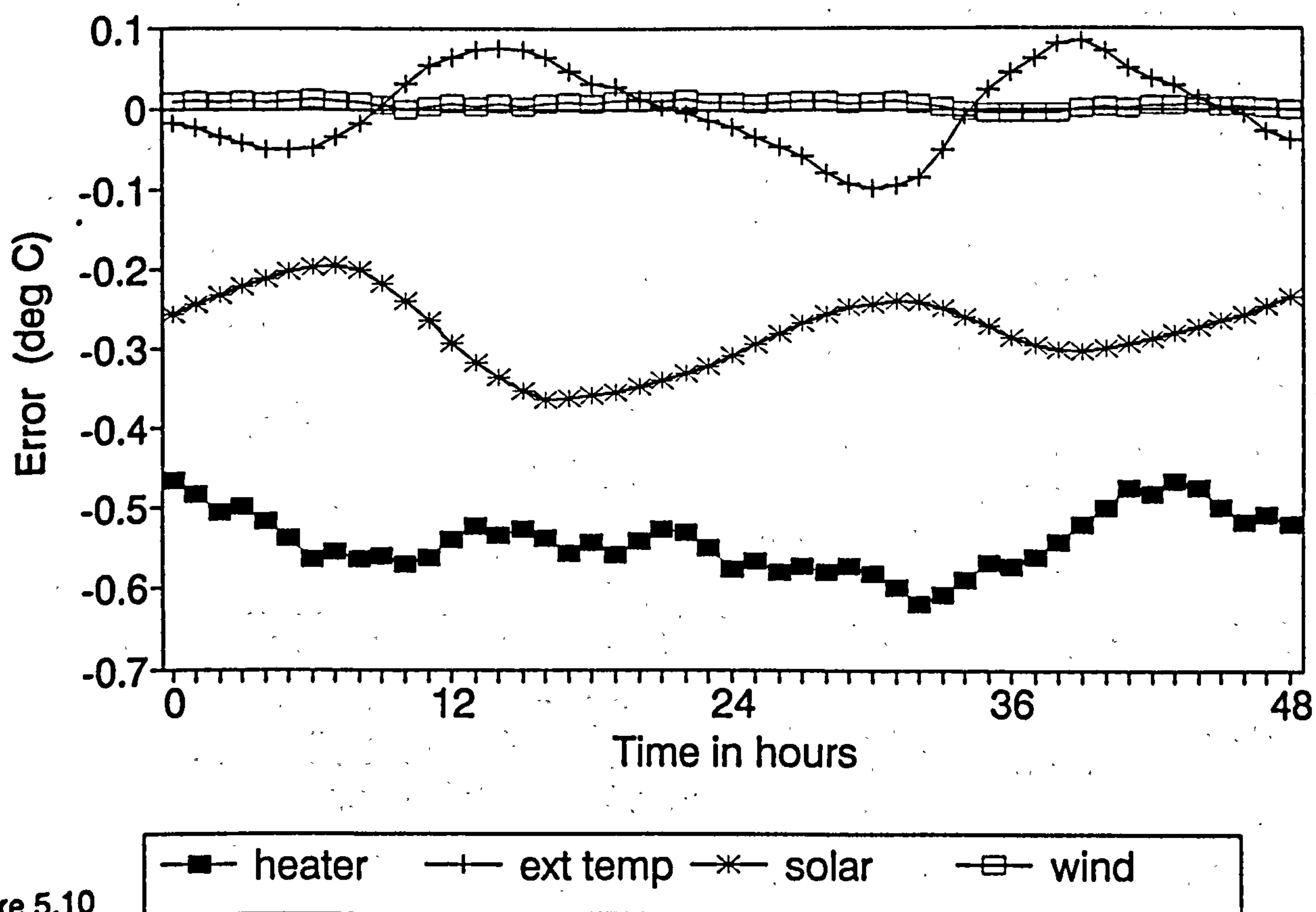
Sensitivity to error in Glazing U-value

Back wall heat flux prediction error responses



Sensitivity to errors in Glazing U-value

Contributions of driving forces to zone temperature error



Sensitivity to errors in Glazing U-value

Reconstituted and observed zone temperature errors

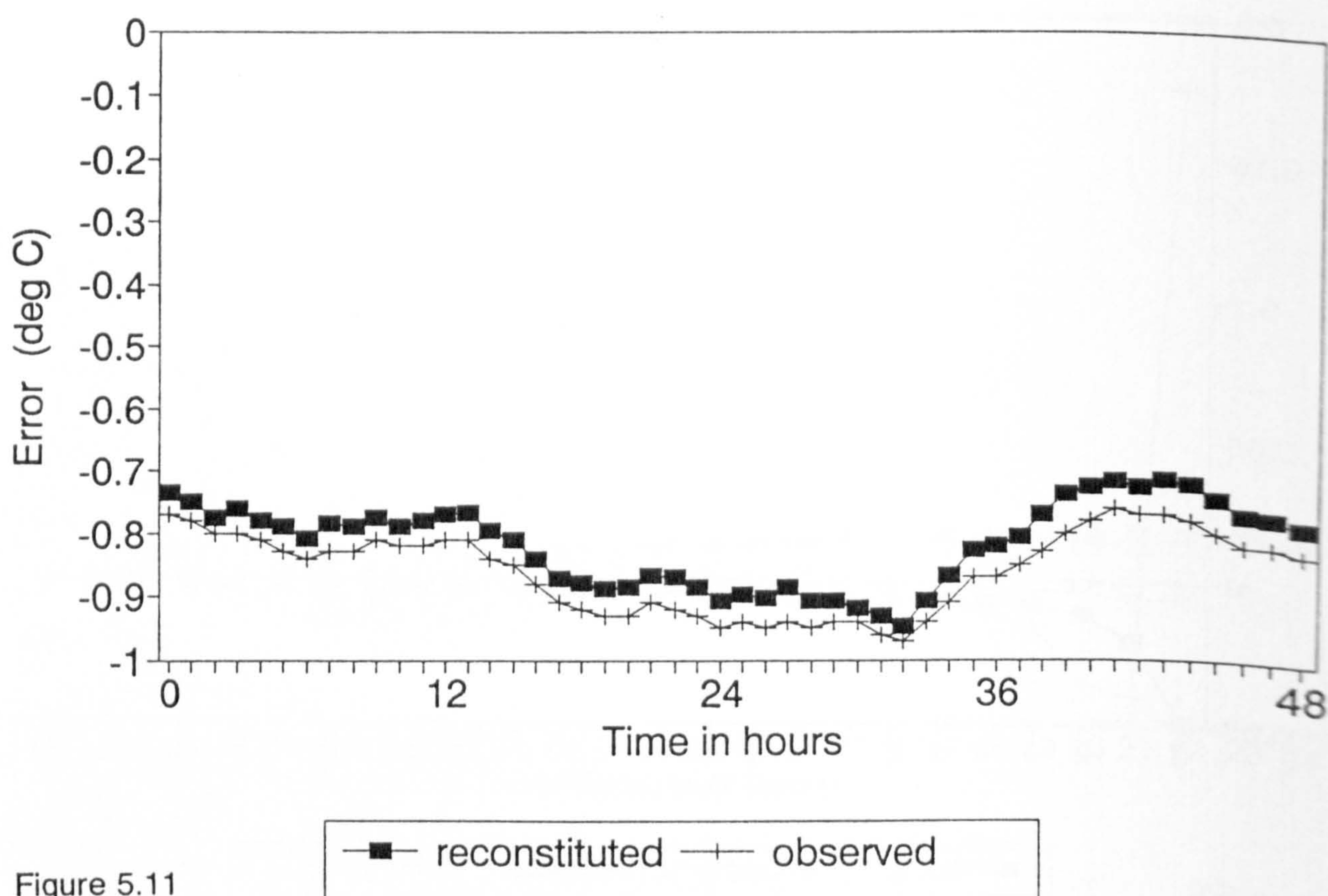


Figure 5.11

Sensitivity to error in Glazing U-value

Relative contributions of driving forces to prediction errors

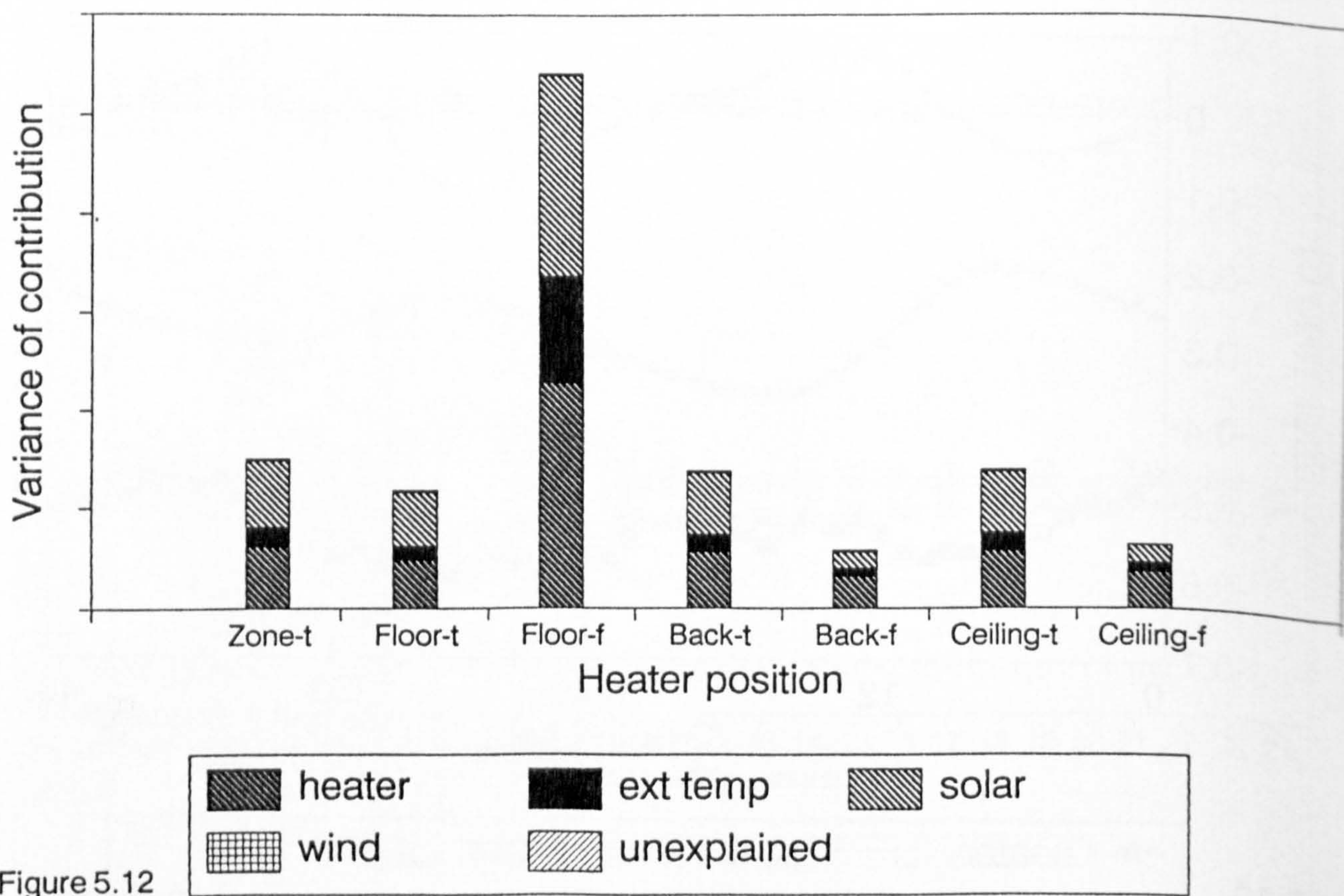


Figure 5.12

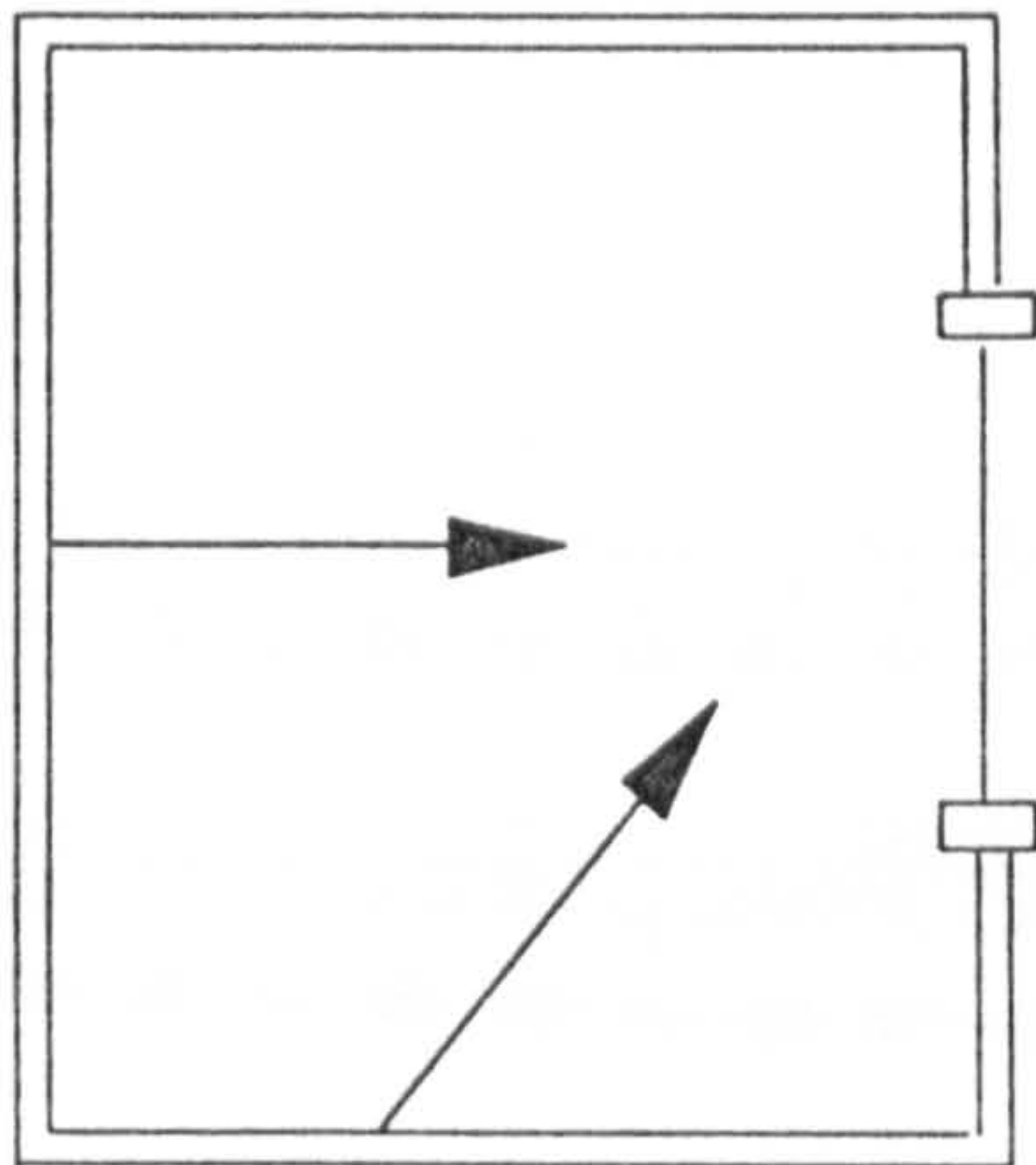


Figure 5.13: Changes in surface heat flux when glazing U-value is increased

Sensitivity to error in Styrofoam conductivity

Zone temperature prediction error responses

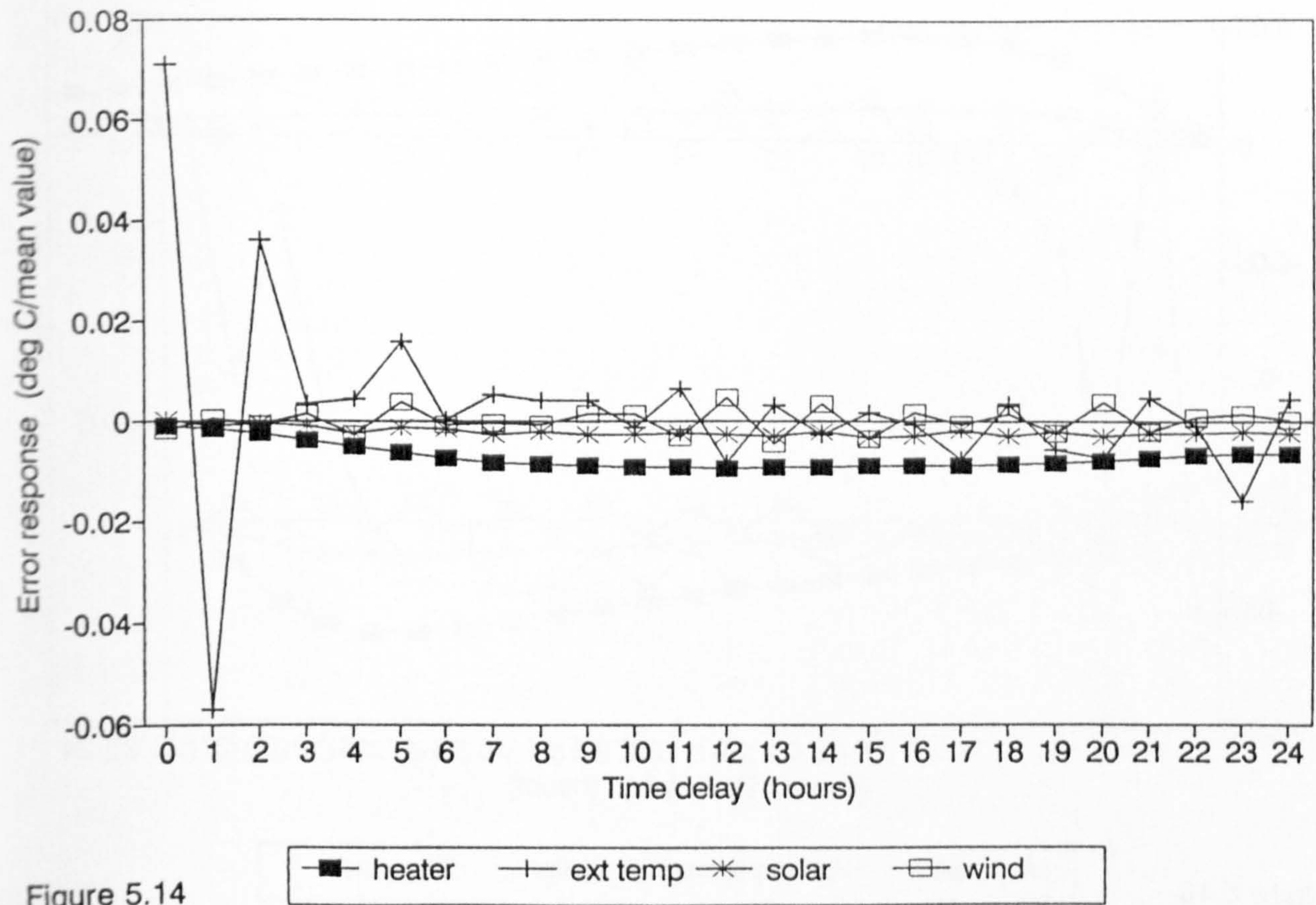
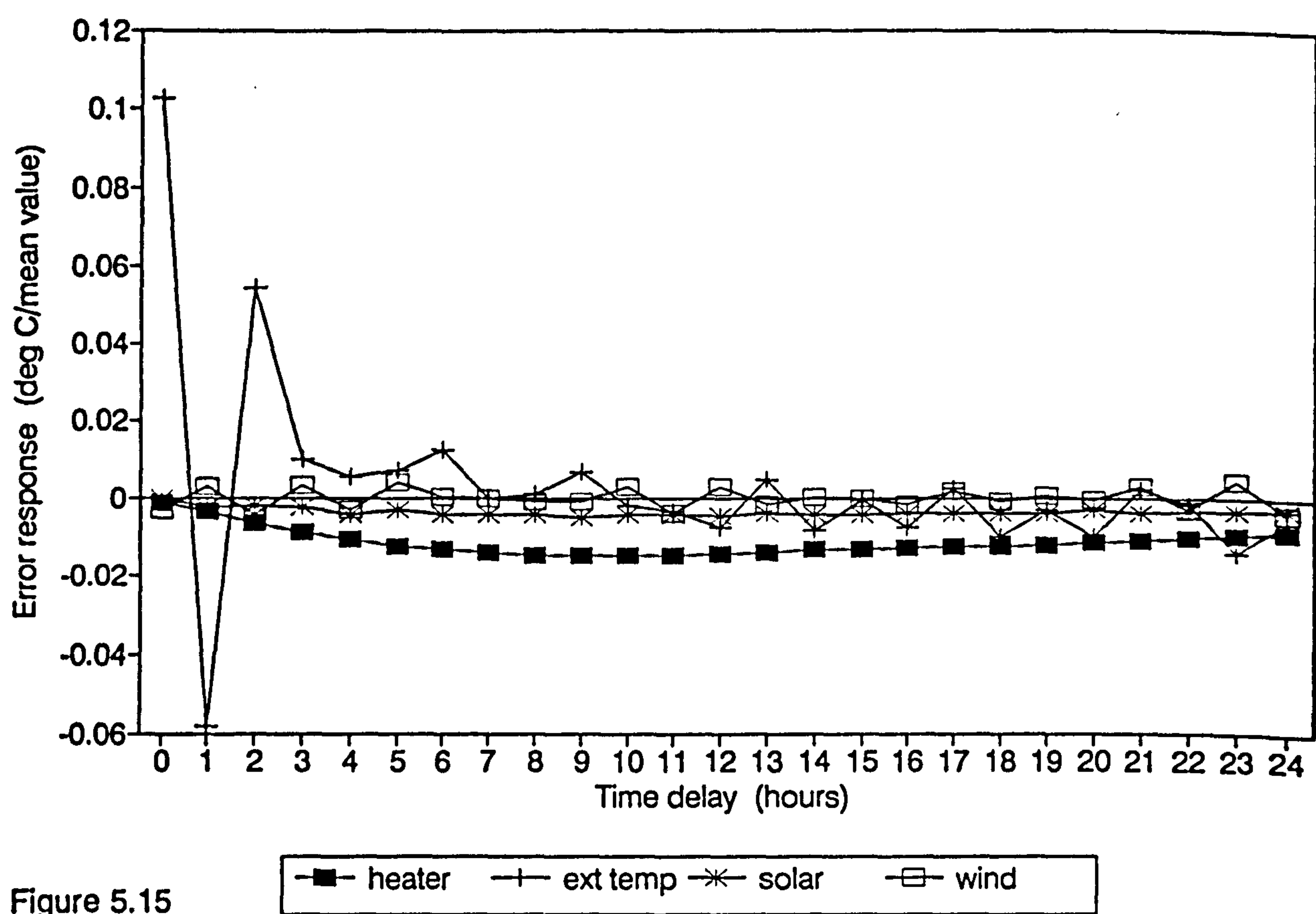


Figure 5.14

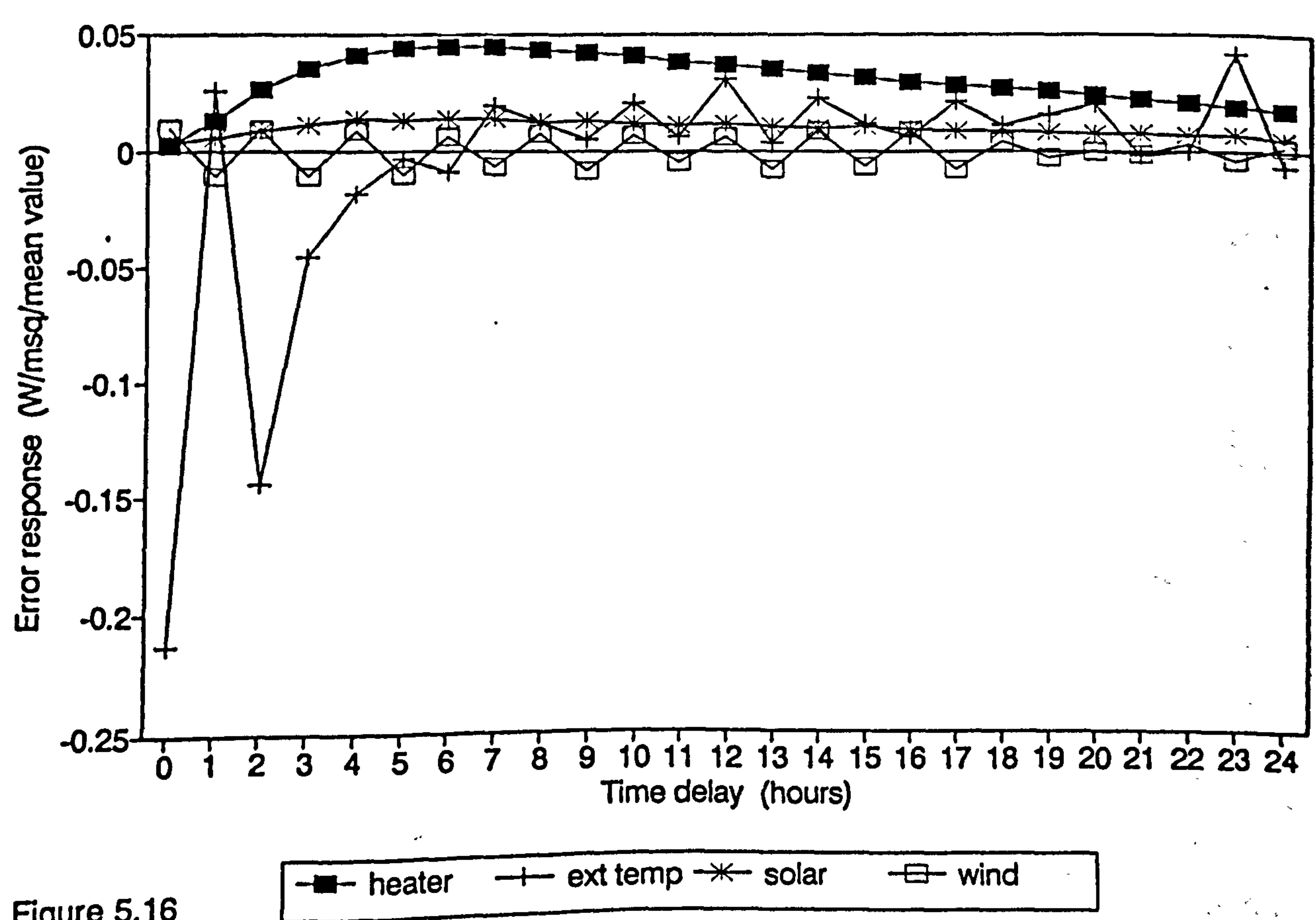
Sensitivity to error in Styrofoam conductivity

Floor temperature prediction error responses



Sensitivity to error in Styrofoam conductivity

Floor heat flux prediction error responses



Sensitivity to error in Styrofoam conductivity

Back wall temperature prediction error responses

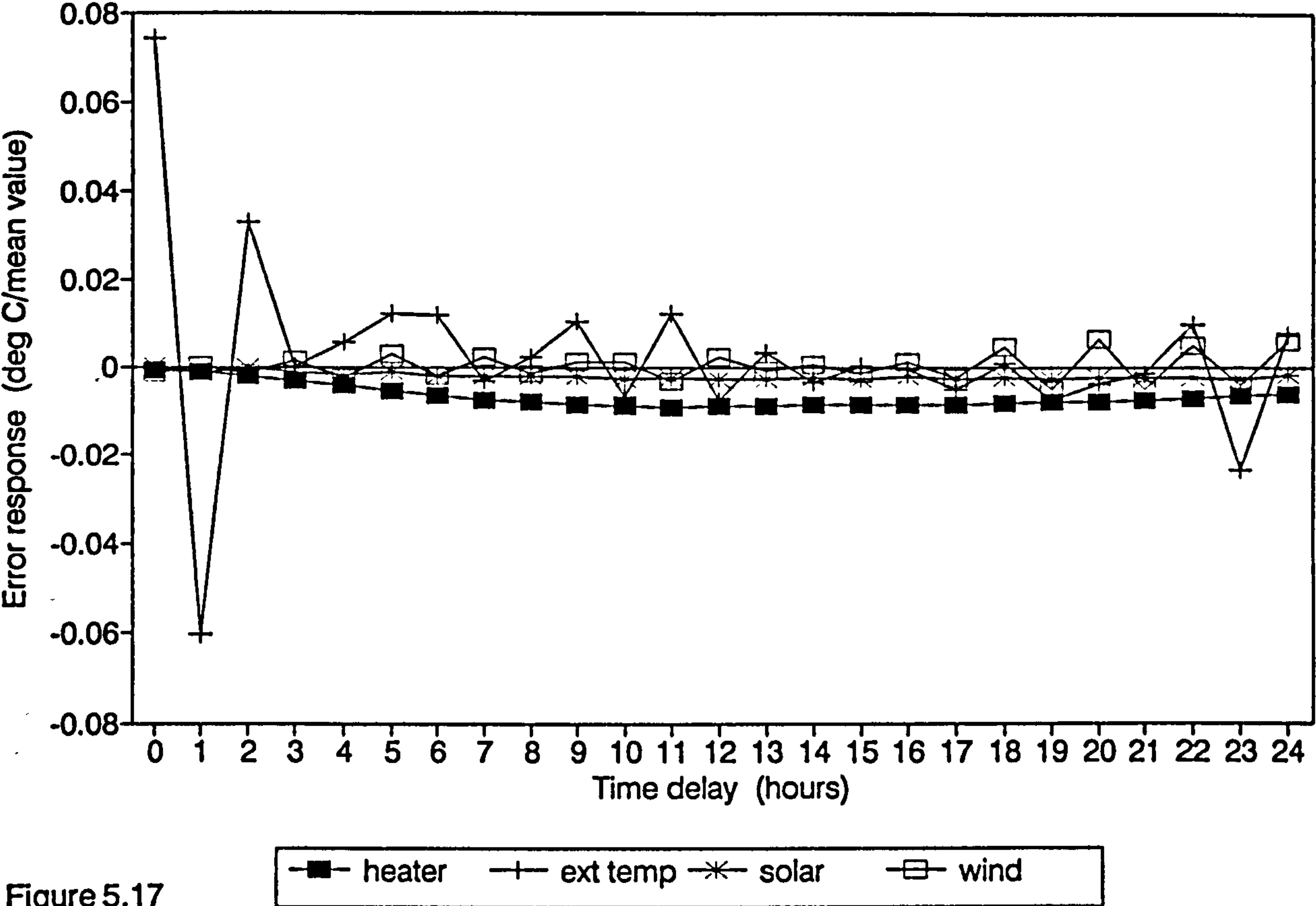


Figure 5.17

Sensitivity to error in Styrofoam conductivity

Back wall heat flux prediction error responses

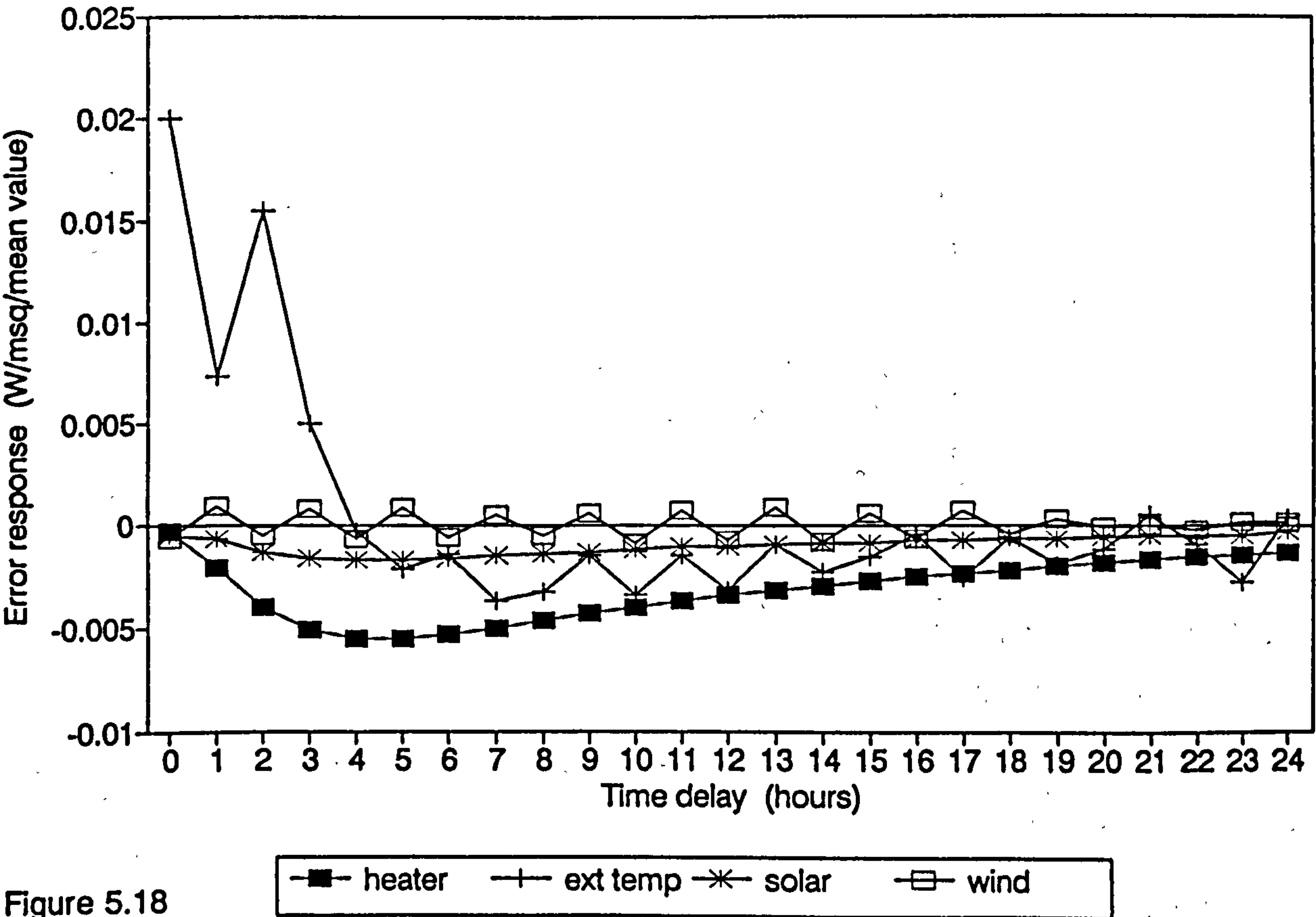


Figure 5.18

Sensitivity to errors in Styro conductivity

Relative contributions of driving forces to prediction errors

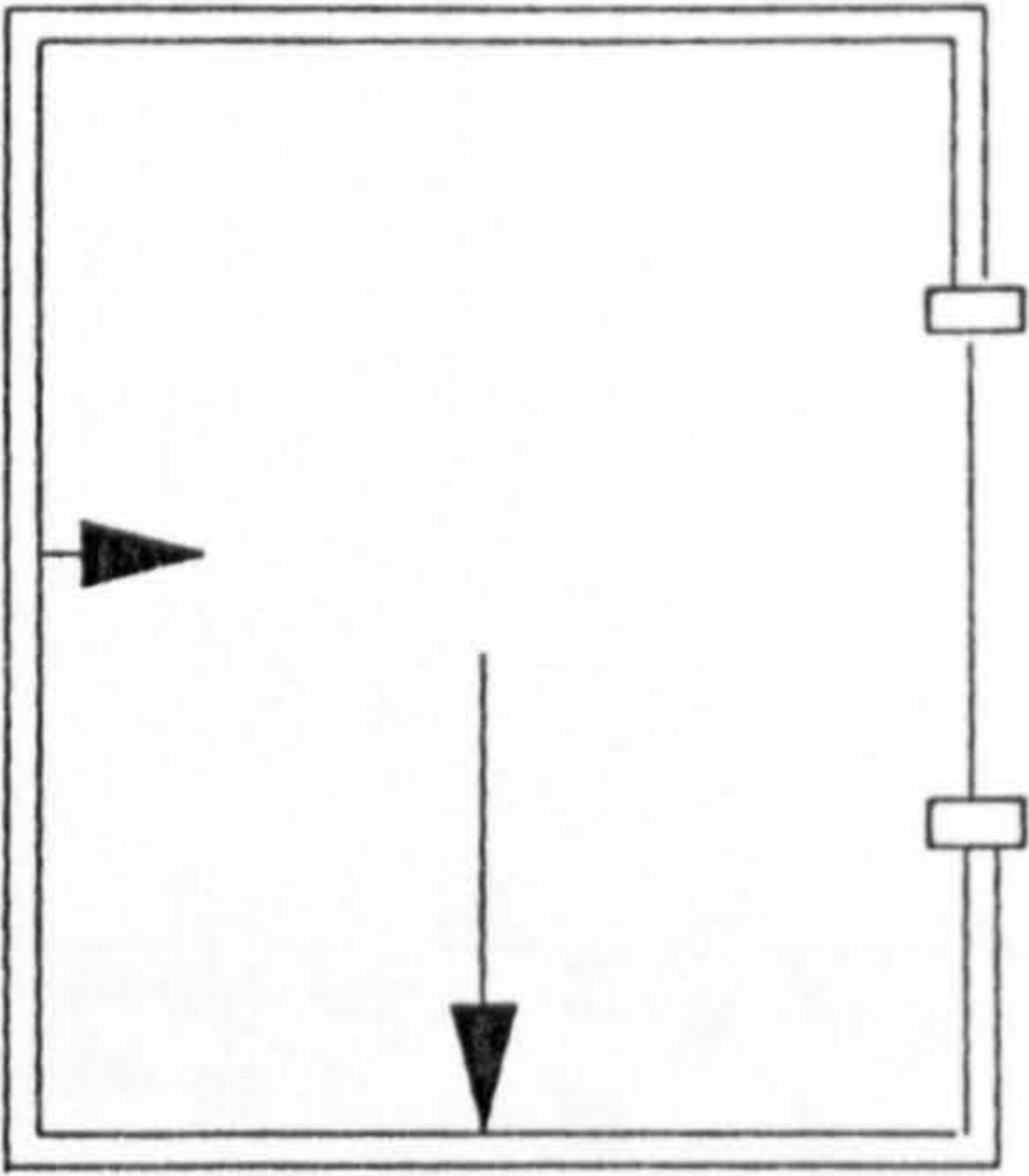
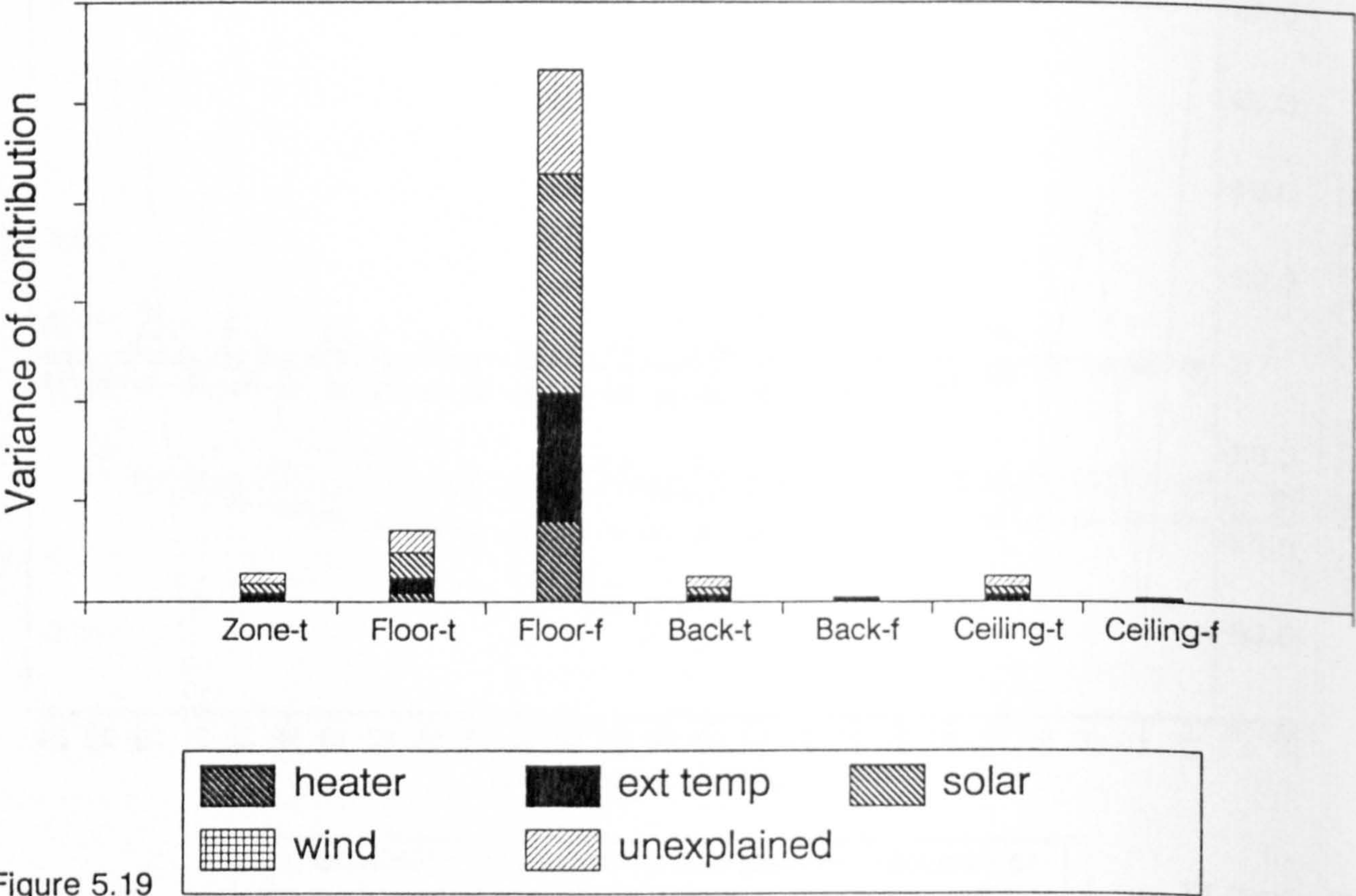


Figure 5.20: Changes in surface heat flux when floor U-value is increased

Sensitivity to error in Solar-lost

Zone temperature prediction error responses

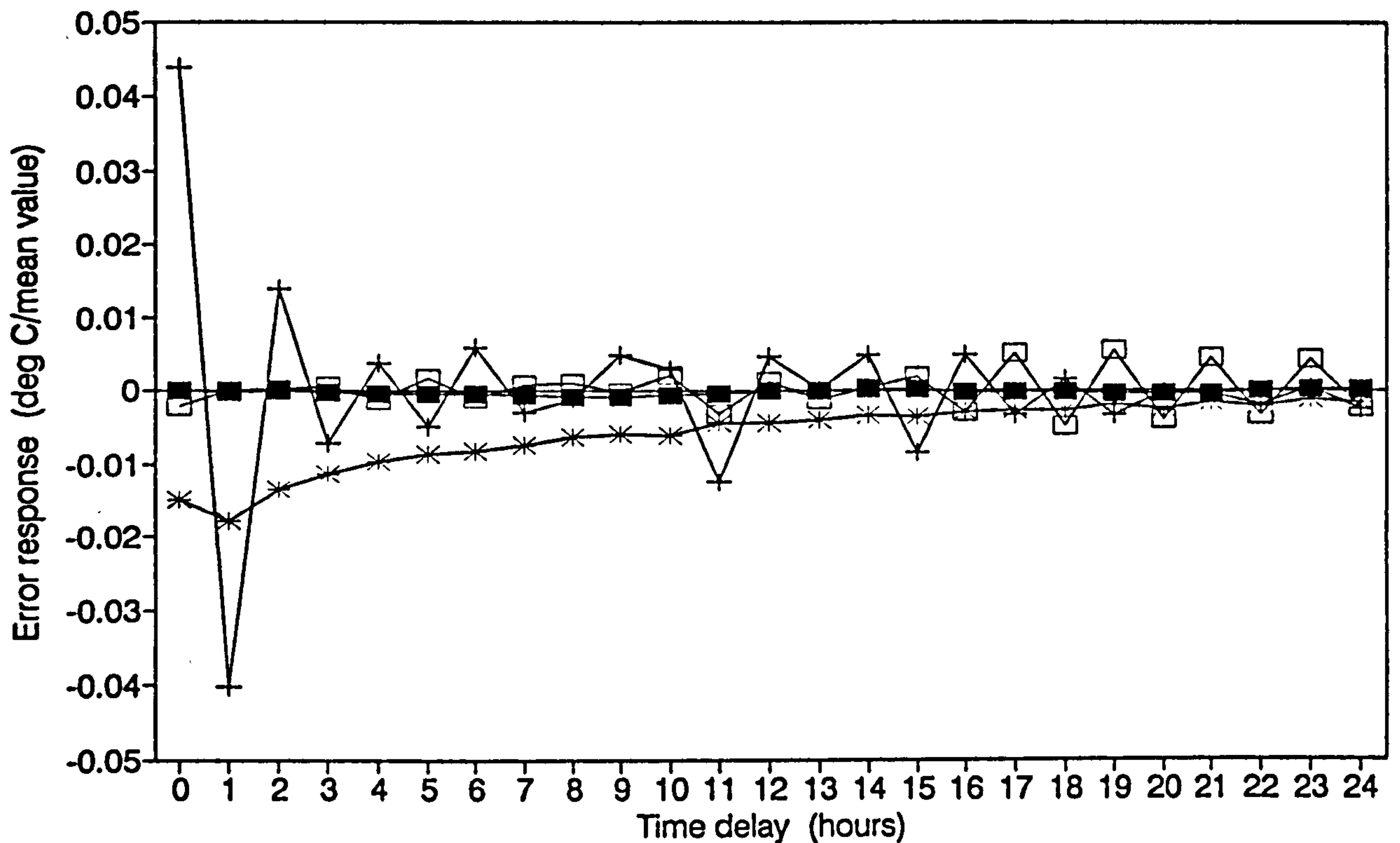


Figure 5.21

Sensitivity to error in Solar-lost

Floor temperature prediction error responses

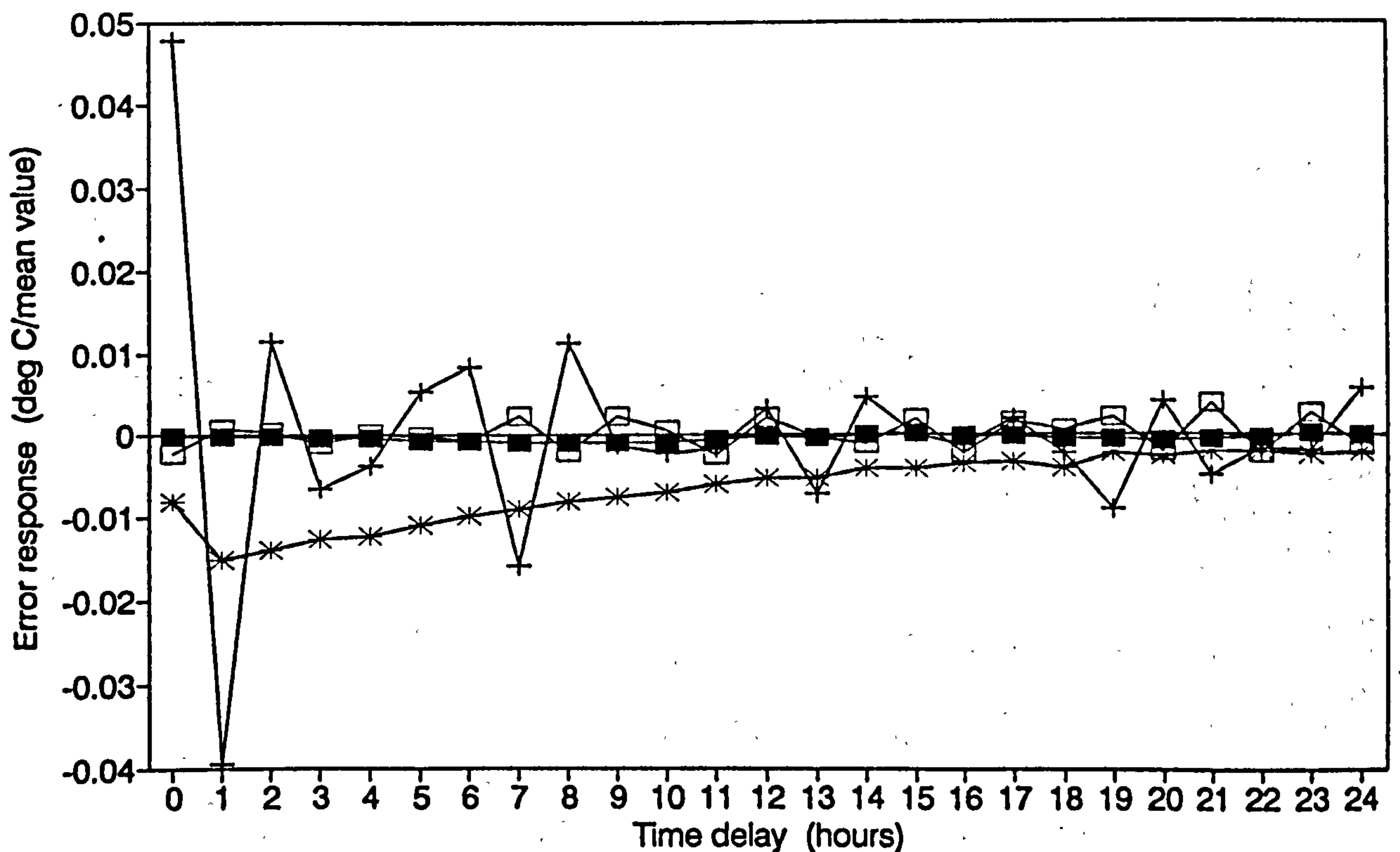


Figure 5.22

Sensitivity to error in Solar-lost Floor heat flux prediction error responses

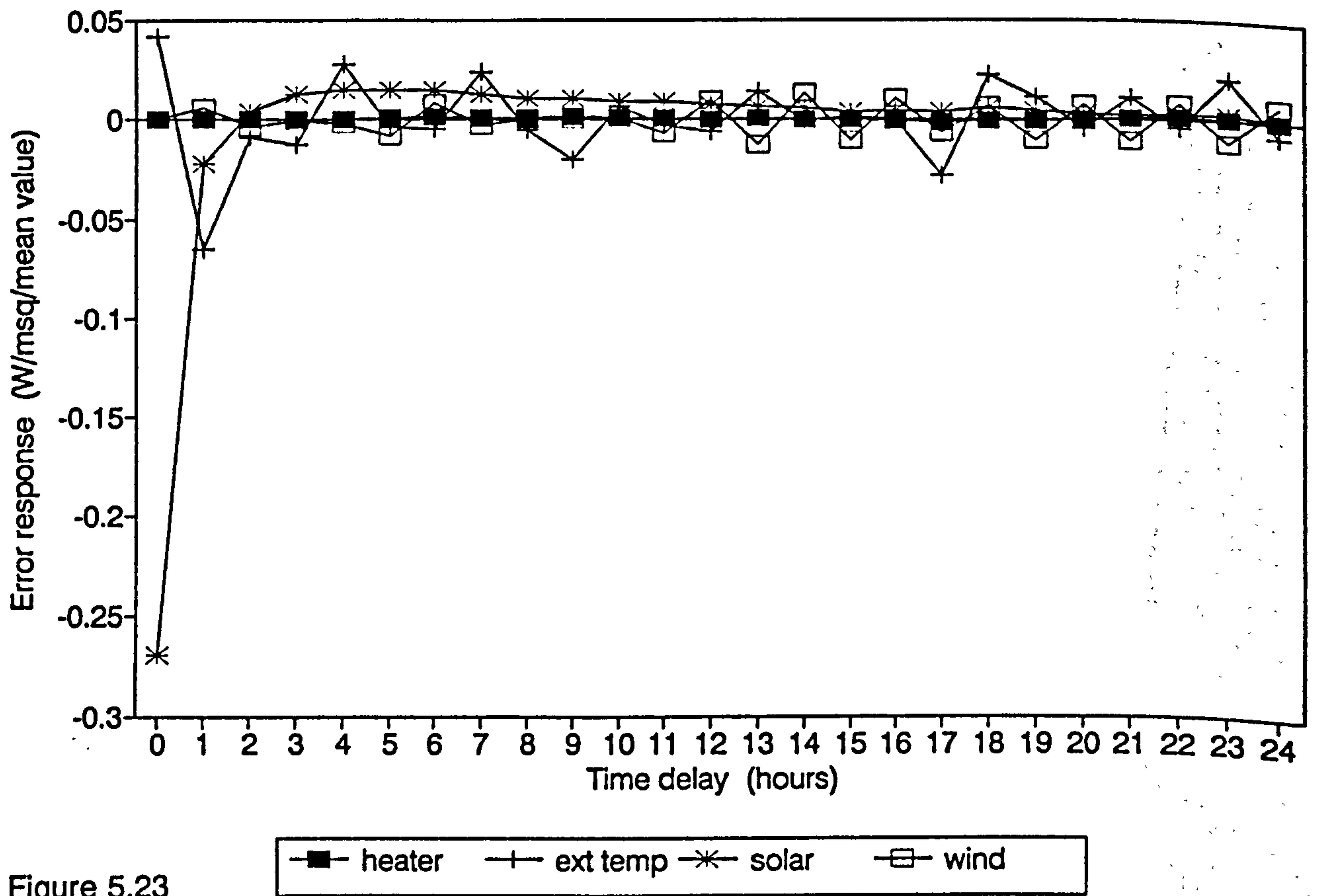


Figure 5.23

Sensitivity to error in Solar-lost Back wall temperature prediction error responses

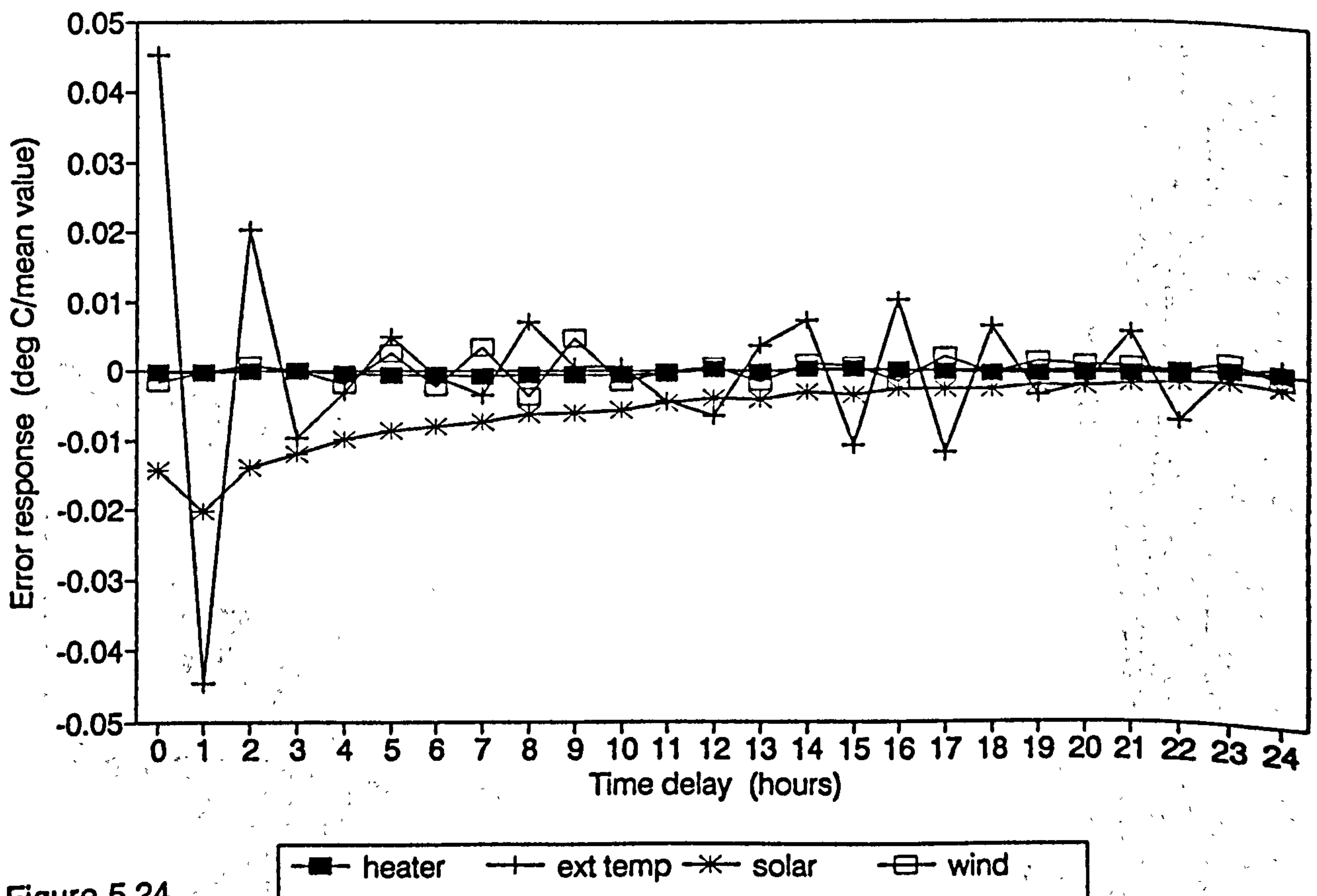
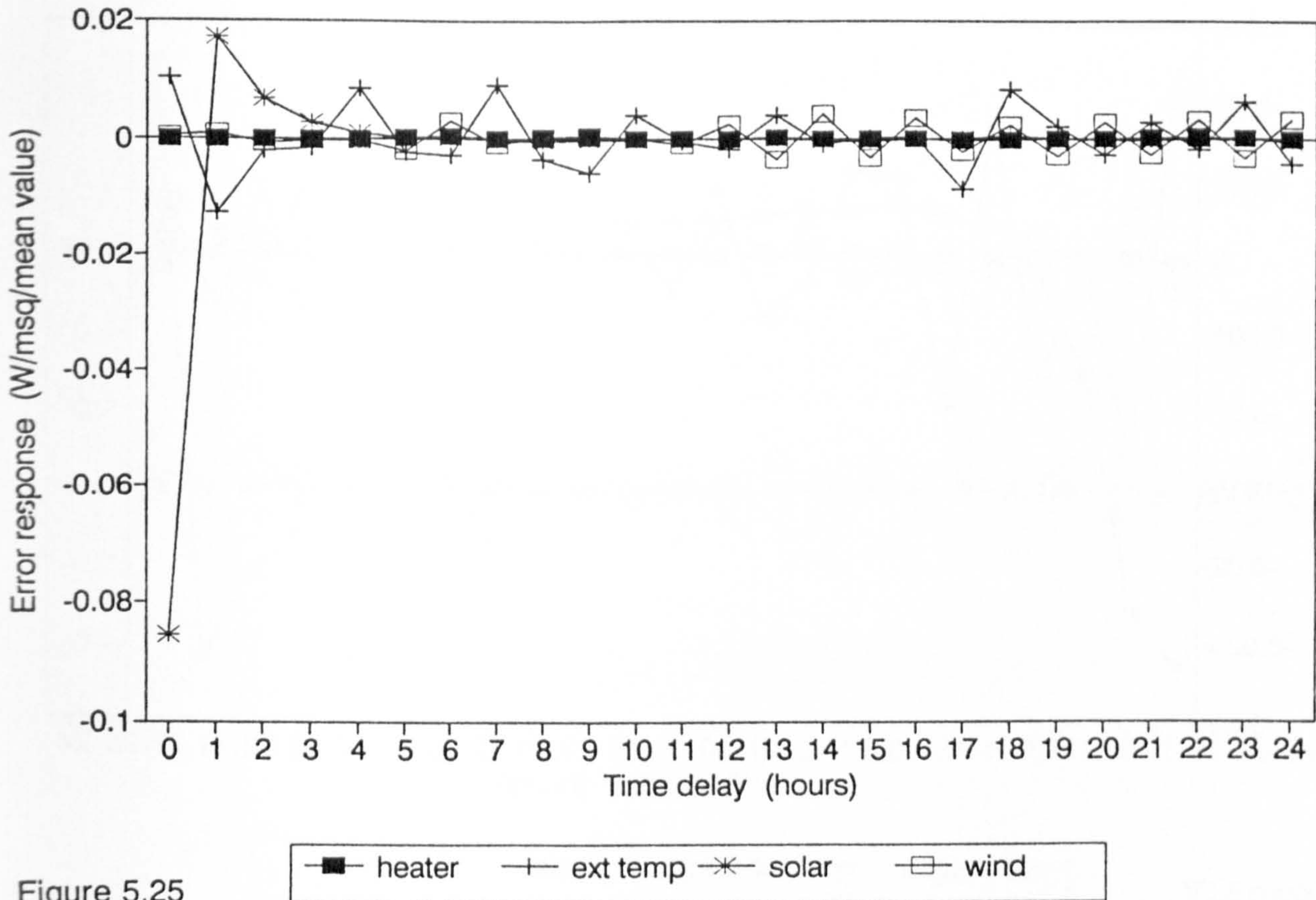


Figure 5.24

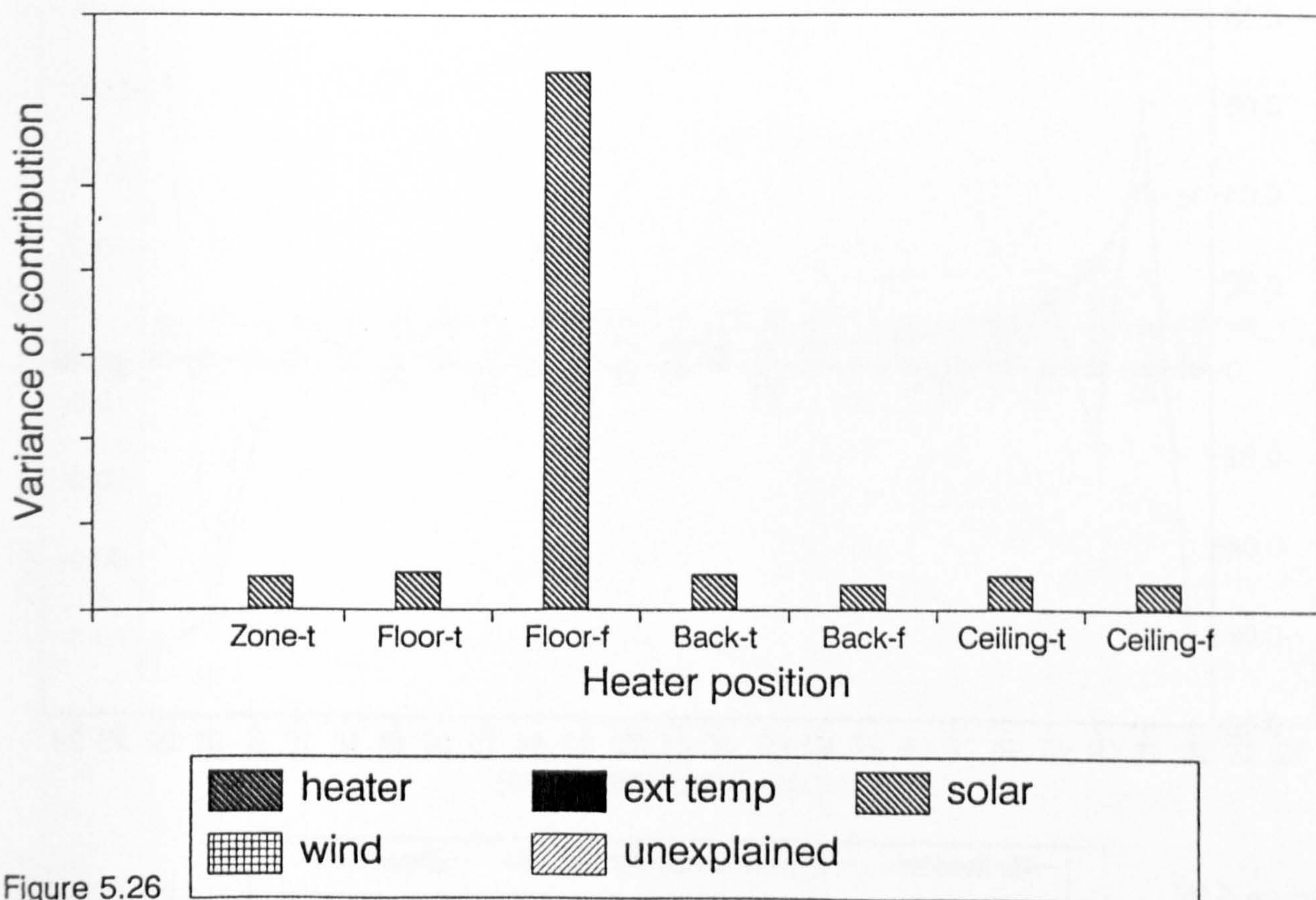
Sensitivity to error in Solar-lost

Back wall heat flux prediction error responses



Sensitivity to errors in Solar-lost

Relative contributions of driving forces to prediction errors



Sensitivity to error in Solar distribution

Zone temperature prediction error responses

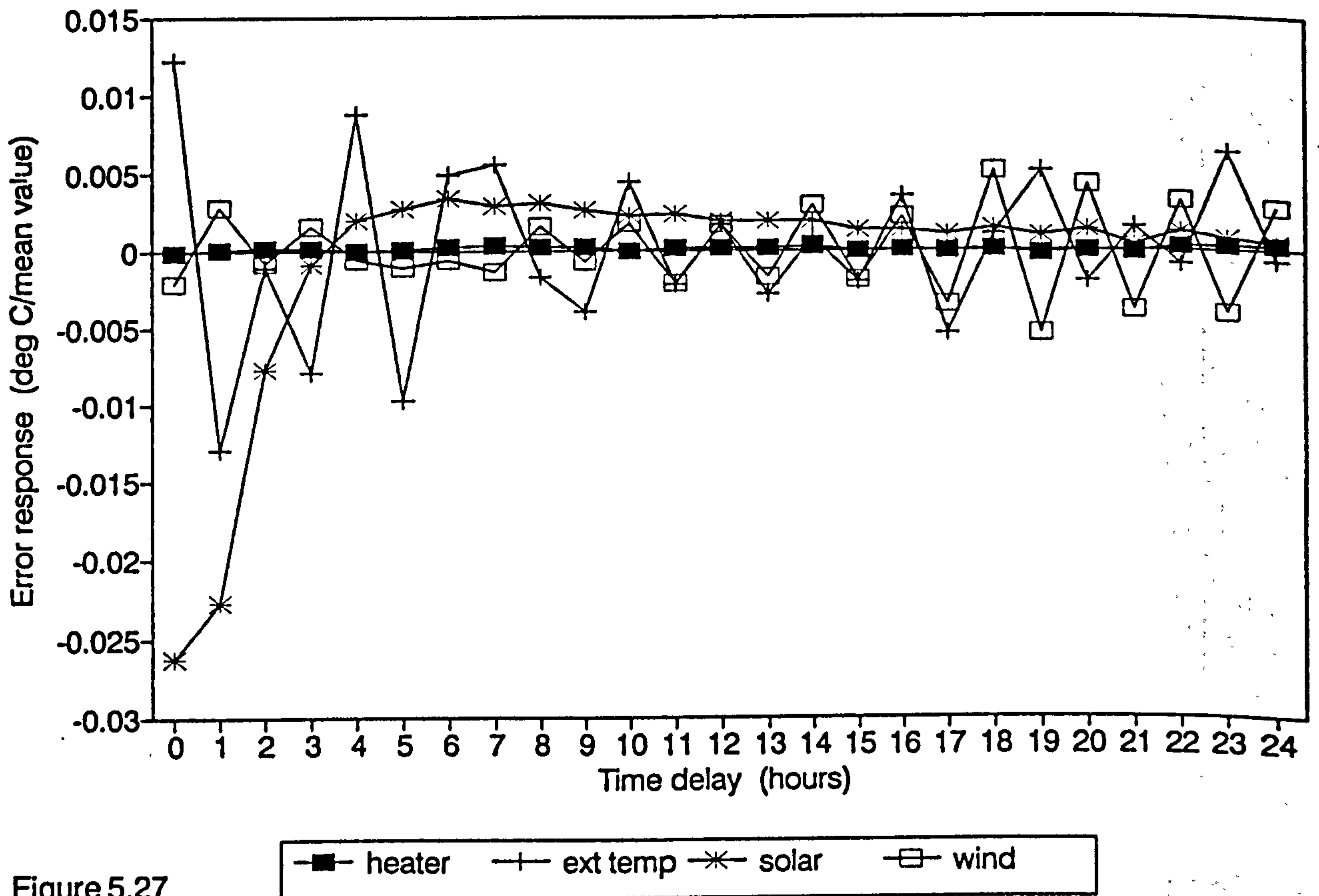


Figure 5.27

Sensitivity to error in Solar distribution

Floor temperature prediction error responses

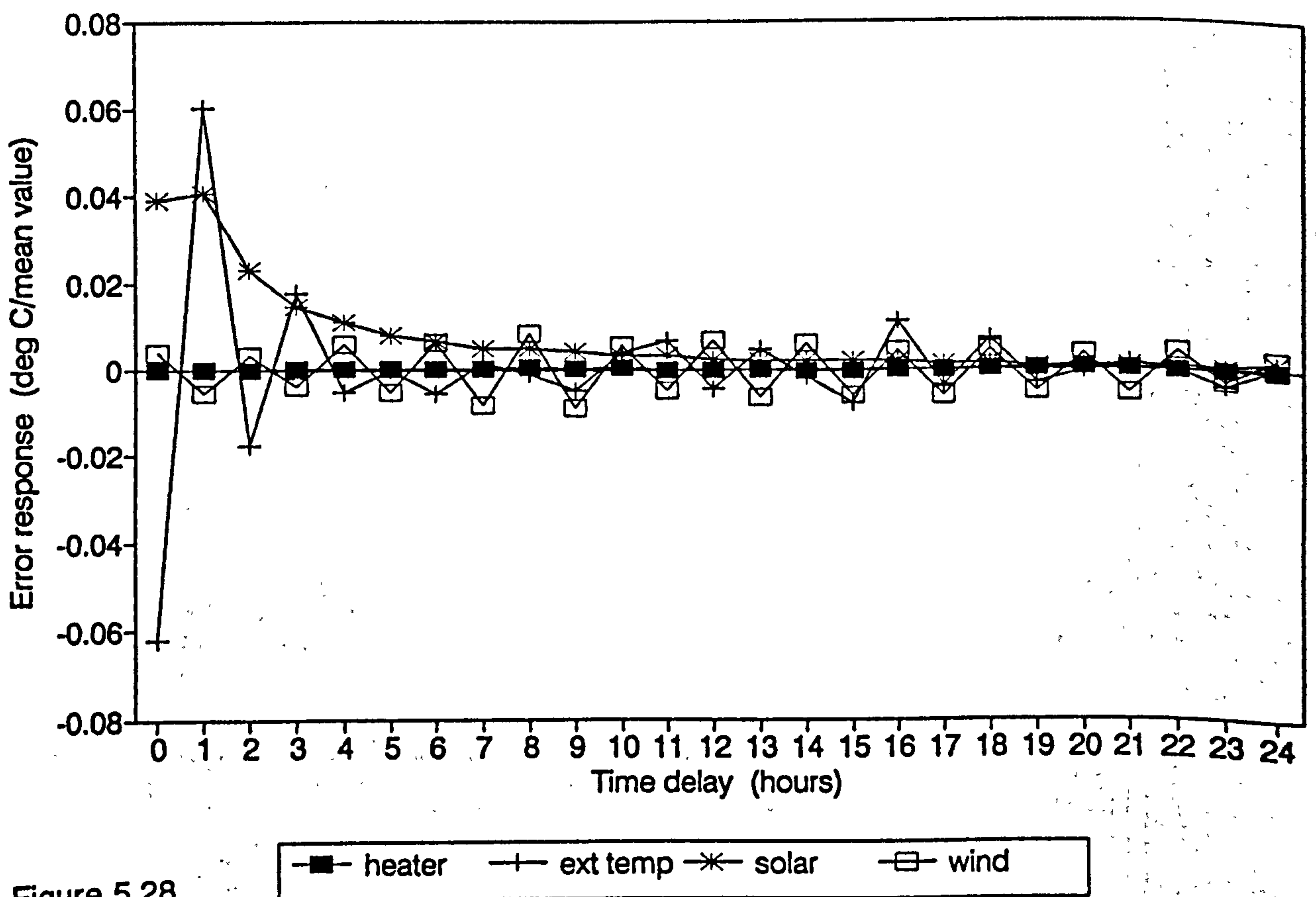
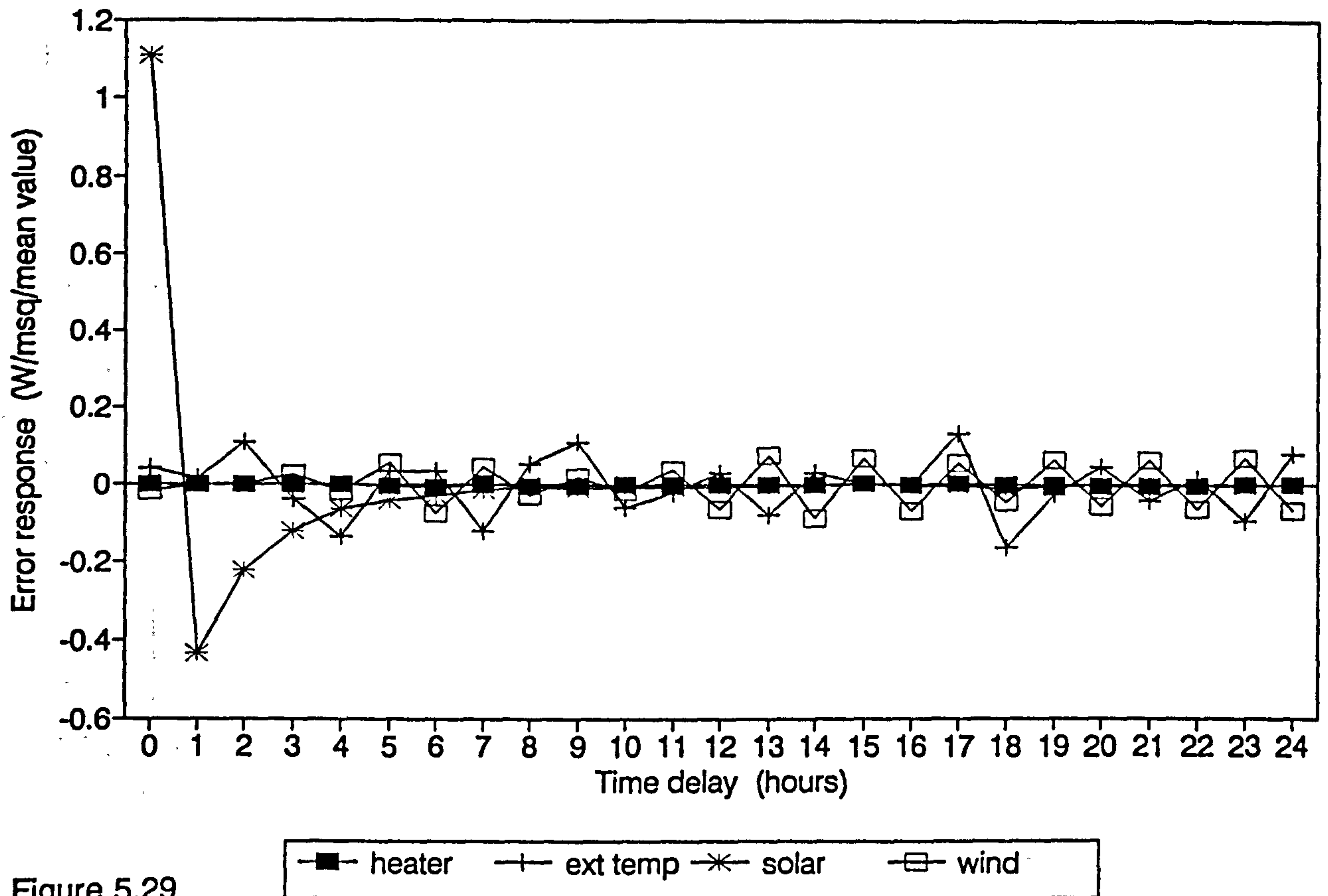


Figure 5.28

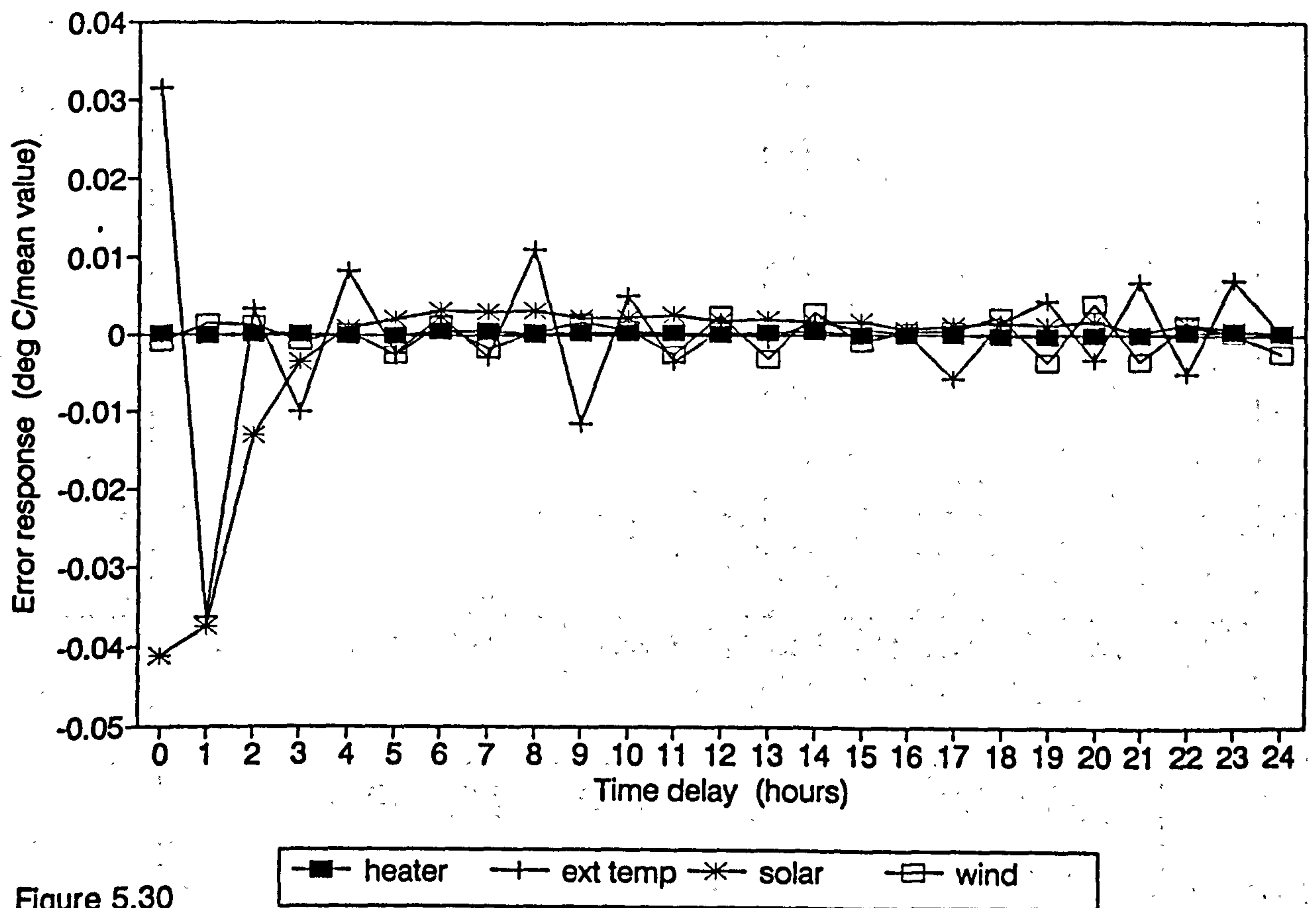
Sensitivity to error in Solar distribution

Floor heat flux prediction error responses



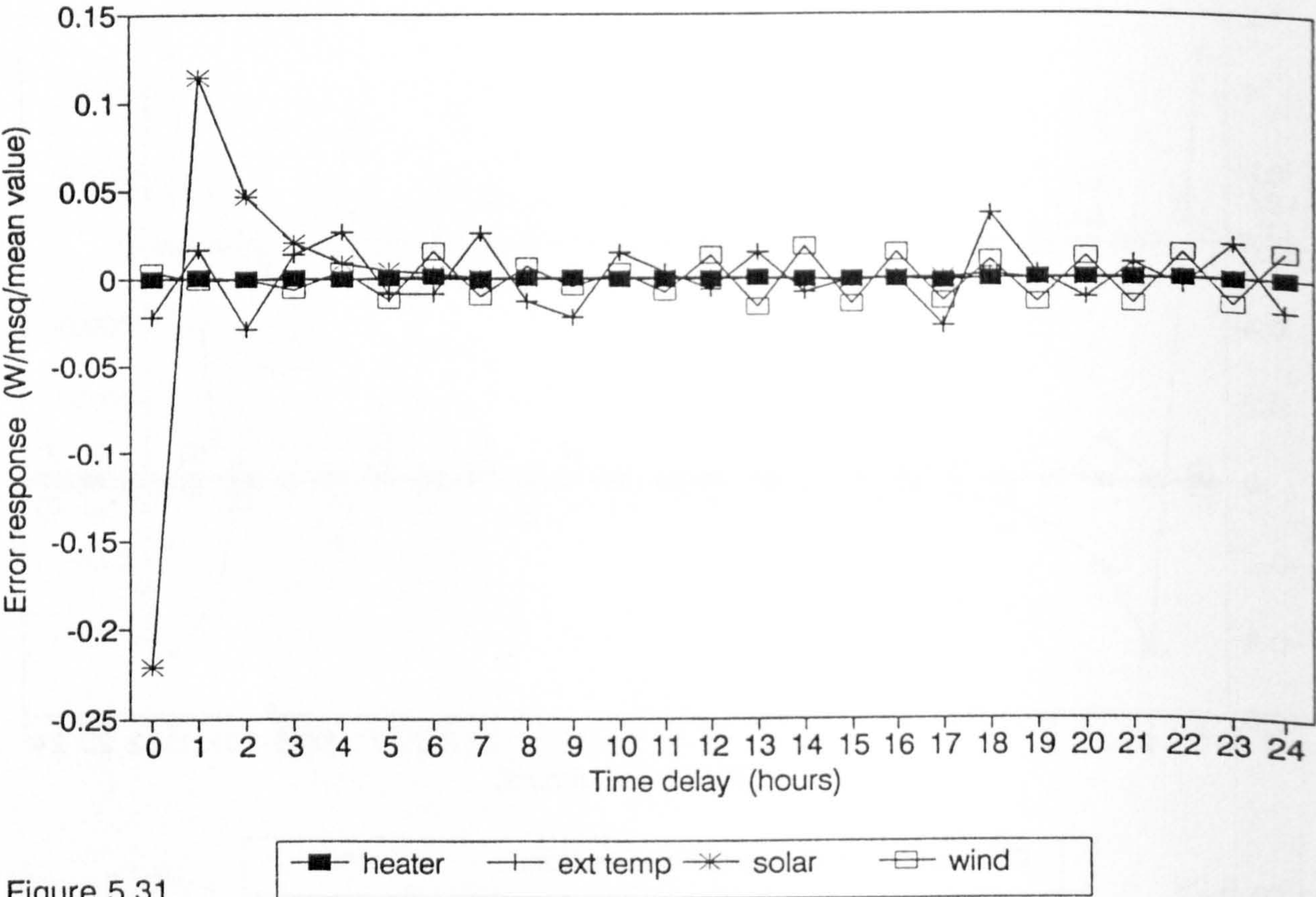
Sensitivity to error in Solar distribution

Back wall temperature prediction error responses



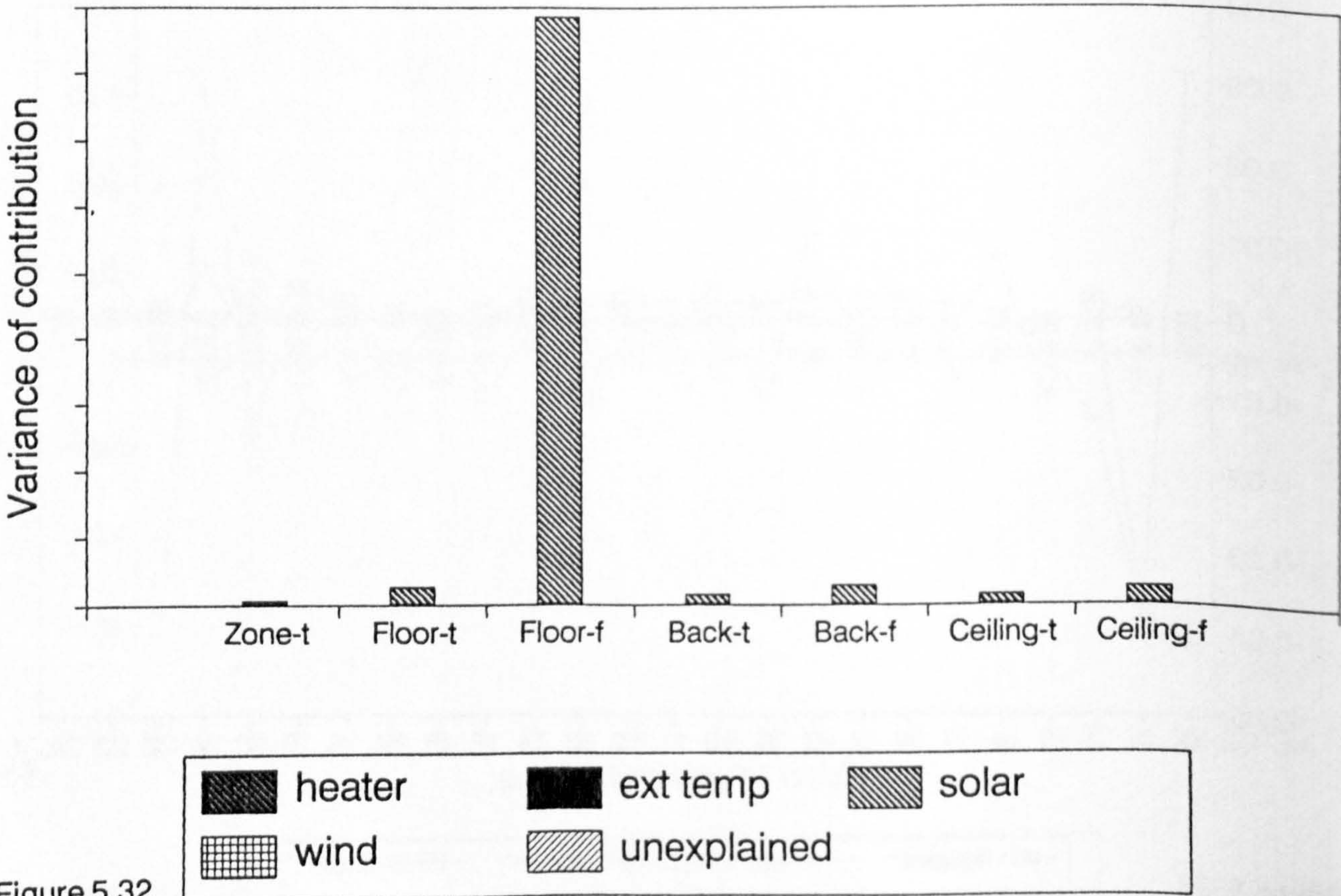
Sensitivity to error in Solar distribution

Back wall heat flux prediction error responses



Sensitivity to errors in Solar distribution

Relative contributions of driving forces to prediction errors



Sensitivity to error in Floorslab density

Zone temperature prediction error responses

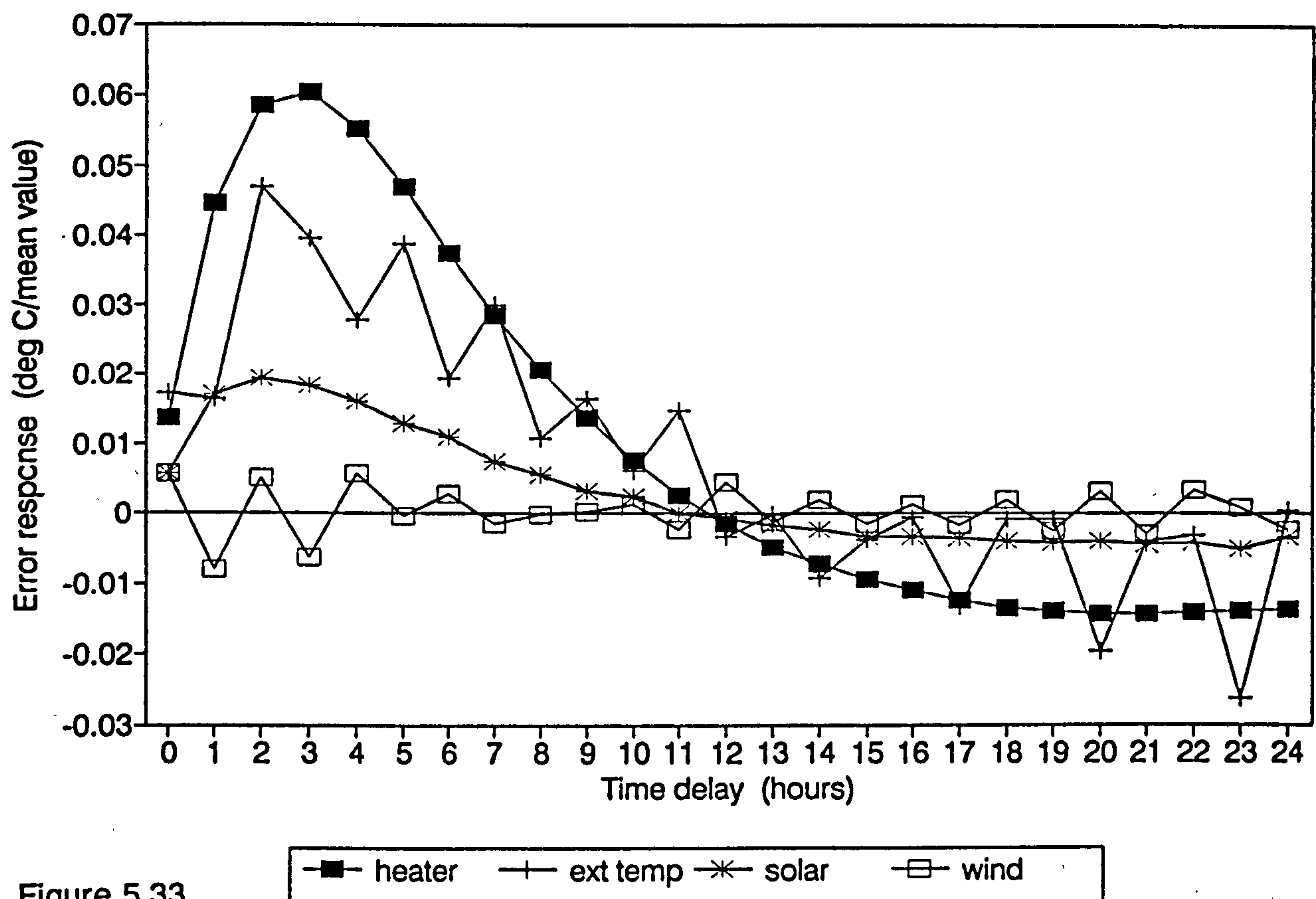


Figure 5.33

Sensitivity to error in Floorslab density

Floor temperature prediction error responses

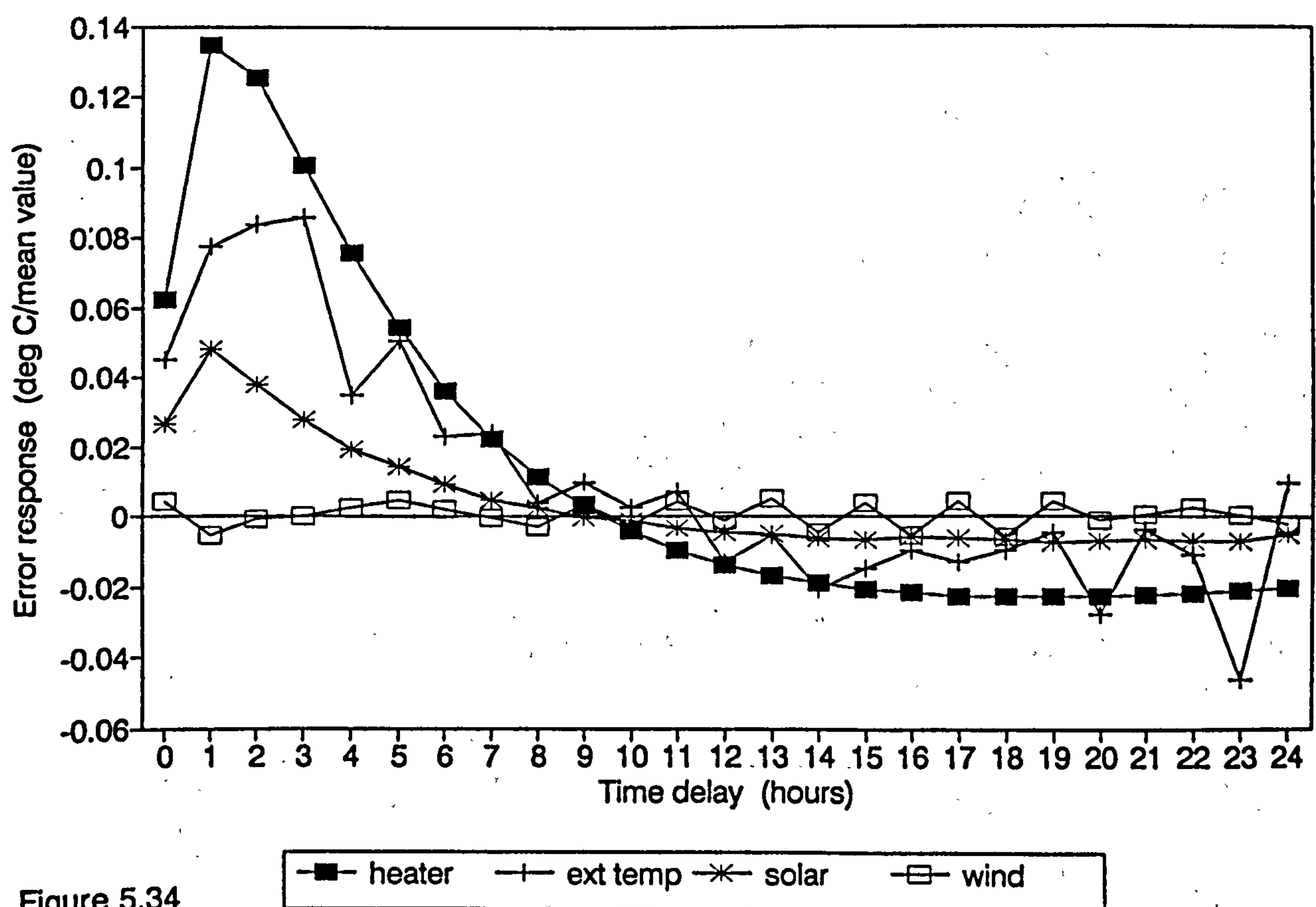


Figure 5.34

Sensitivity to error in Floorslab density

Floor heat flux prediction error responses

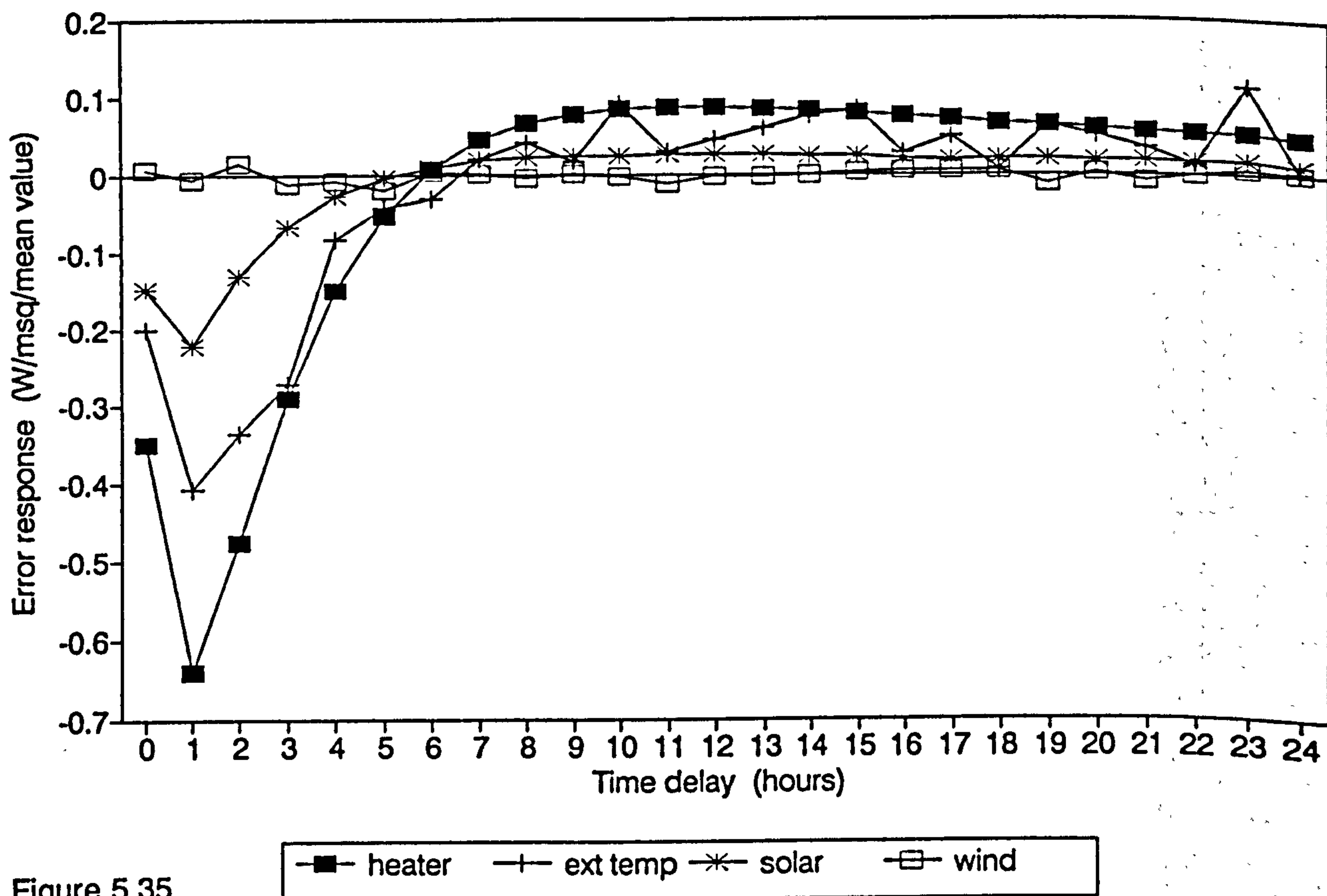


Figure 5.35

Sensitivity to error in Floorslab density

Back wall temperature prediction error responses

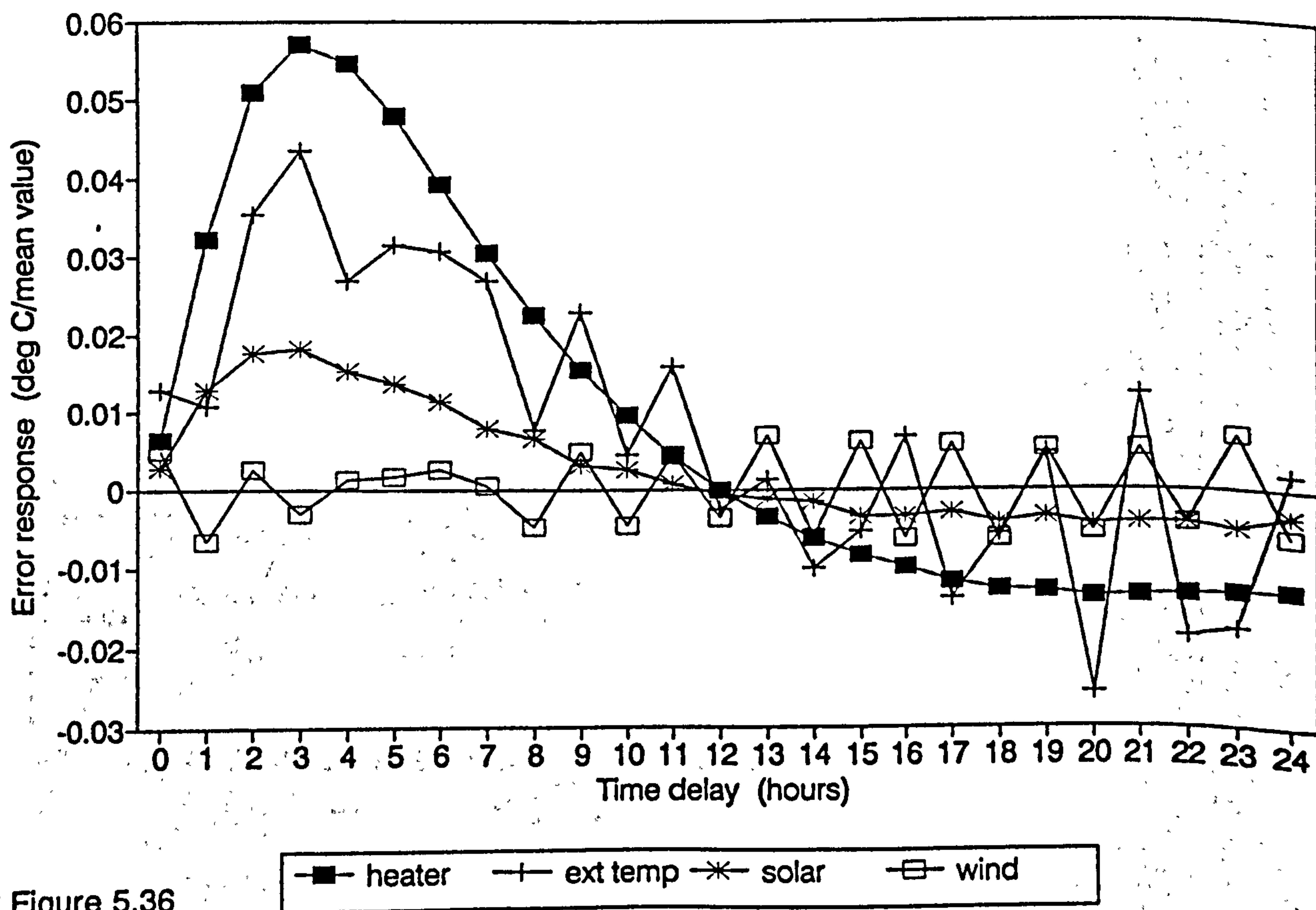


Figure 5.36

Sensitivity to error in Floorslab density

Back wall heat flux prediction error responses

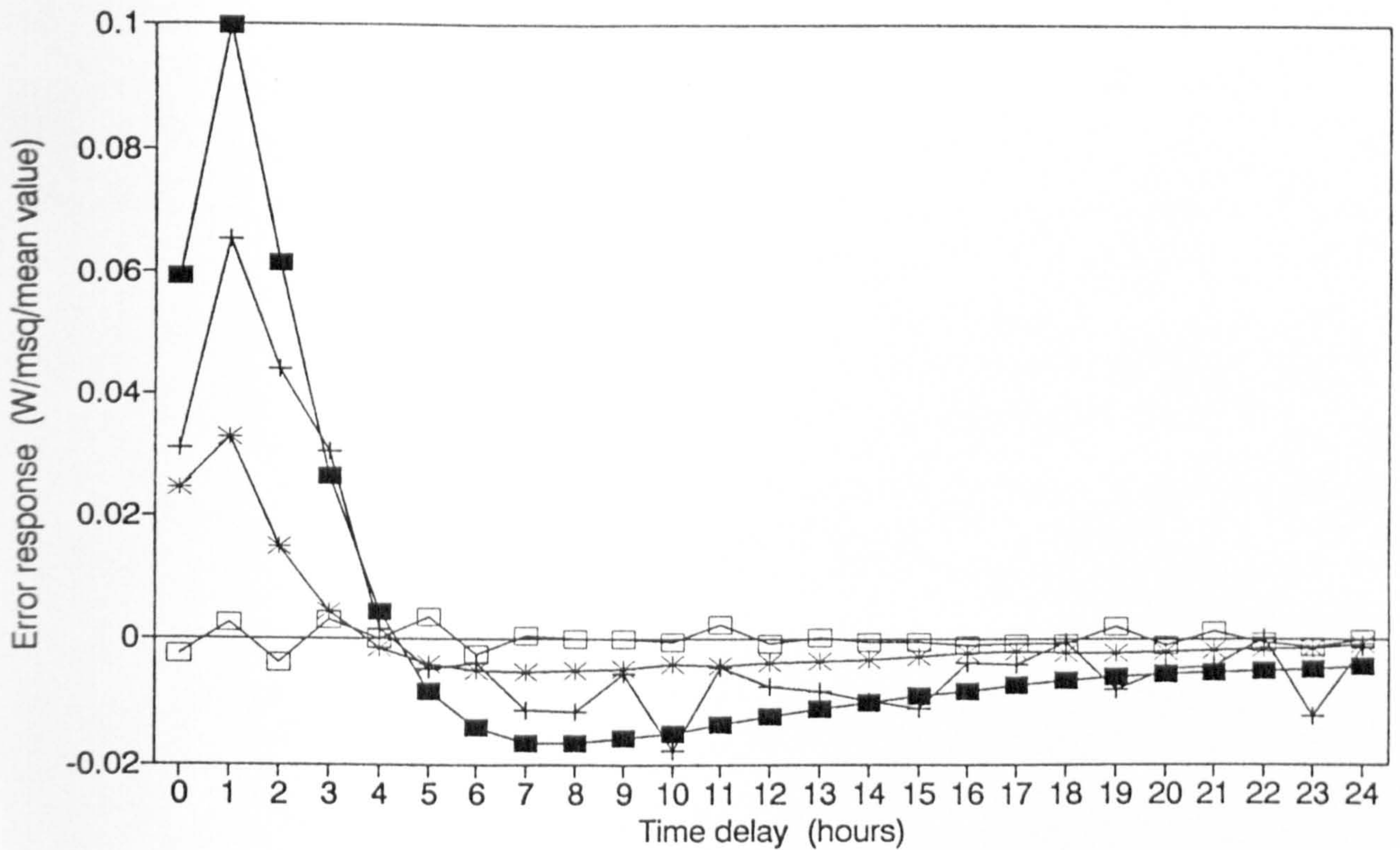


Figure 5.37

Sensitivity to errors in floorslab density

Relative contributions of driving forces to prediction errors

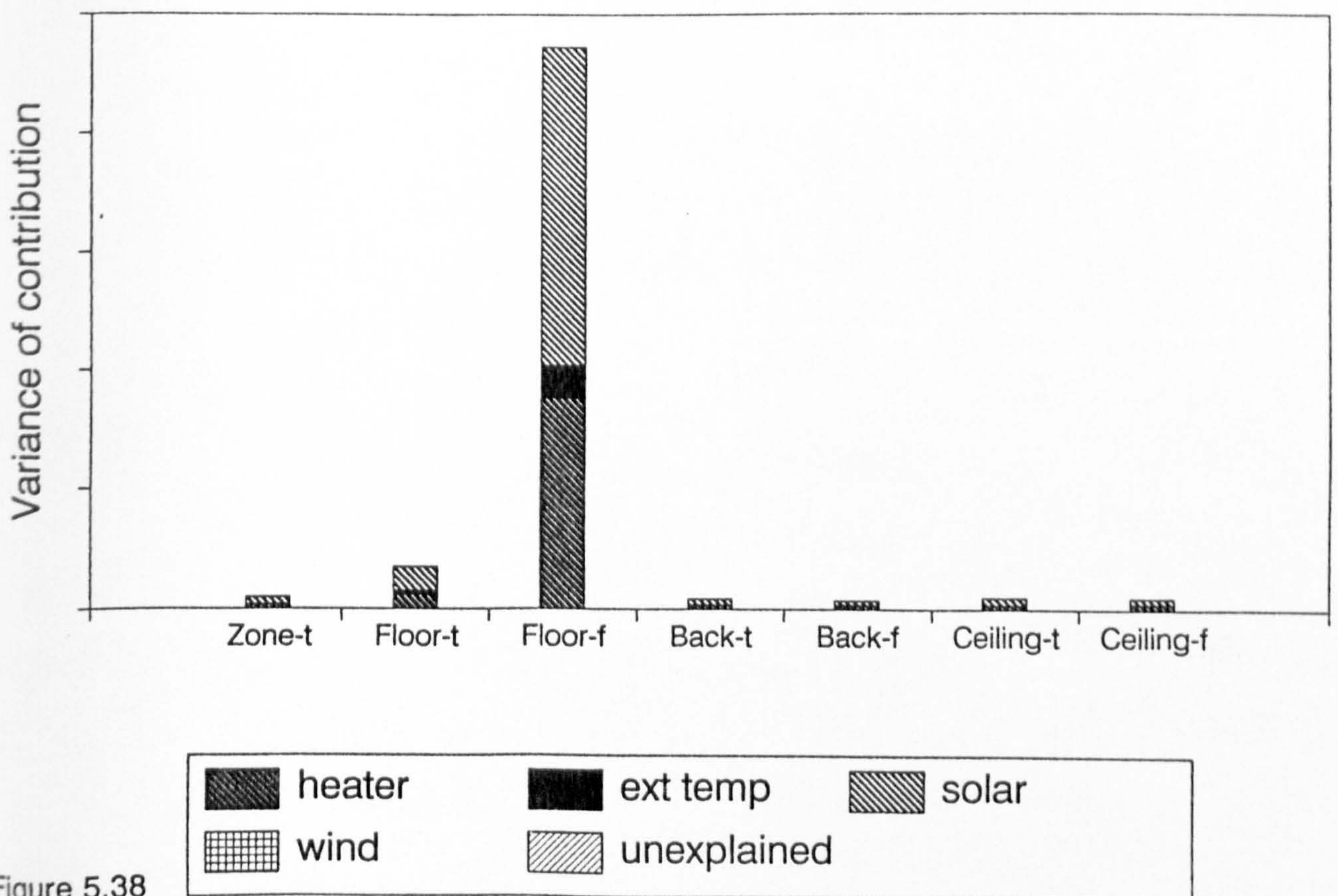


Figure 5.38

CHAPTER 6

VALIDATION USING REAL DATA

In this chapter the new technique is applied to a real empirical validation problem. The design of the experiment is first described. This consists of the choice of a suitable test building, the choice of the most appropriate way in which to operate it, and the operation of the simulation model under test.

The resulting dataset is then compared with the predictions of the model SERI-RES. The deconvolution technique is used to analyse the discrepancies between simulations and measurements in two rooms with alternative heater positions. In this way the appropriateness of the new technique for its intended purpose will be demonstrated.

6.1 Choice of buildings for validation

Early attempts at the empirical validation of thermal simulation models concentrated on the use of data from complex, often occupied buildings. The reason for this is plain: the aim of any validation exercise is to test, and hopefully prove, a model in a situation as close as possible to that in which it will ultimately be applied.

All too often, these early attempts at validation failed to produce conclusive results. A recent review [Lomas, 1985] concluded that, out of 600 available datasets, only 27 were from experiments with sufficient control of uncertainties to allow meaningful conclusions to be drawn when they had been used. This was partly due to the lack of a consistent approach to gathering data and conducting simulations, and partly due to the lack of a consistent criterion for what actually constituted validity. However, the underlying problem with these early experiments was the very high degree of uncertainty which inevitably resulted when modelling such complex, occupied buildings.

A logical response to this problem was to attempt to eliminate at least some of the sources of uncertainty from validation experiments. Data can, in principle, be gathered from a whole range of test facilities. Test houses contain all the constructional features of the real buildings with which simulation models will ultimately be used, but they eliminate the large uncertainties due

to the behaviour of occupants. Individual test rooms of realistic construction eliminate the zone structure of real buildings. Optionally, they may also eliminate other features such as ground floor heat loss or ventilation. Simpler test cells can be used when the uncertainties due to realistic construction techniques are to be avoided. Window test calorimeters no longer feature a realistic room geometry. Hot boxes provide a way of testing the performance of single building elements in the laboratory, where they are no longer exposed to real climate. Finally, laboratory tests of material properties provide data on the behaviour of materials from samples which may not even be of realistic size. This progression of complexity is shown graphically on Figure 6.1.

Table 6.1 summarises the elements of complexity which are present in building simulation problems, and indicates which are present in various types of test facility. The table broadly confirms the conclusions of Figure 6.1, that test rooms provide a middle line between excessive complexity and over-simplification, but, more importantly, it also demonstrates the extreme flexibility of test cells, in that many of the mechanisms present in real buildings may be incorporated if required, or alternatively may be disabled. Inevitably, there is a compromise between the degree of realism in an experiment, and the uncertainty which will be encountered when the results are interpreted. Outdoor test rooms can provide a highly flexible vehicle in which to carry out empirical validation experiments, in which degrees of complexity can be progressively introduced as more and more confidence in the capabilities of simulation programmes is gained.

	Laboratory tests	Hot box	Test room	Test houses	Occupied houses
Size		●	●	●	●
Climate			●	●	●
Enclosure geometry			●	●	●
Construction			○	●	●
Ground floor heat loss			○	●	●
Natural ventilation			○	○	●
Interzonal heat transfer			○	○	●
Casual gains			○	○	●
Occupancy					●

Table 6.1: Features available in alternative test facilities

(● denotes always present ○ denotes present if required)

6.1.1 The Energy Monitoring Company Outdoor Test Room facility

In 1985 the Energy Monitoring Company (EMC) established an outdoor facility consisting of eight test rooms. A ninth room was added in 1986, and a tenth in 1993. The original eight rooms are built in semi-detached pairs. These are timber framed buildings, with an external skin of painted plywood. The walls and ceiling of each room is insulated with glass fibre quilt between the timber framing members, and are lined with plasterboard. The floor is insulated with solid Styrofoam, and a layer of concrete slabs on top of this insulation provides further thermal mass. They are tightly sealed, and this sealing is periodically checked by pressurisation testing. If required, ventilation can be introduced in a controlled fashion by a mechanical ventilation system. The buildings feature interchangeable south facades, which allow different areas of glazing to be quickly installed. Figure 6.2 shows a section through one of the rooms.

Over their lifetime the rooms have been periodically 'matched': that is experiments have been carried out to ensure that pairs of rooms perform identically when they are identically configured.

The rooms are highly monitored. The basic instrumentation consists of 20 temperature measurements in each room, measurement of the electrical power used for heating and the volume of air used by the mechanical ventilation system. A small number of surface heat flux sensors are also available. Data acquisition is controlled by an IBM-compatible personal computer, which also controls the heating to the rooms and the operation of the mechanical ventilation systems. The output of each sensor is sampled at a rate determined by the speed of response of the sensor, which ranges from 10 seconds for solar radiation measurements to five minutes for temperature sensors. The resulting data is averaged and periodically recorded on the hard disk of the data acquisition computer. The data used here was recorded at five minute intervals, and subsequently reduced to hourly average values. The test rooms and monitoring equipment are described in detail in [Martin and Watson, 1989].

6.2 Operating buildings for model validation

Early attempts to carry out empirical model validation using data from the EMC test rooms revealed that early data sets collected in the rooms were not ideally suited to the task [Martin and Watson 1991]. Two principal problems were identified when the data was analysed:

- cross-correlation analysis of the resulting simulation errors failed to identify correctly the source of those errors. The reason for this was found to be correlations between the forces driving the rooms: solar radiation, external temperature and auxiliary energy input.
- in the interests of realism the data sets, which had been gathered as part of a previous project, featured thermostatic control of the rooms to a fixed setpoint for part of the day. In other words, a control device was used to vary the power input to the rooms, to achieve and then maintain a given temperature within each room. When the corresponding simulations were carried out the heater operating schedule was supplied to the model in the form of a setpoint as a function of time of day.

Clearly there is nothing that can be done about correlations between solar radiation and external temperature - these variables are beyond the experimenters' control. In designing test room experiments to collect data for the purposes of model validation, a number of aspects of room

operation can be controlled. The most obvious such mechanism is the auxiliary heater schedule. It may at least be possible to choose a heater operation strategy which will reduce, or perhaps even eliminate, correlations between this driving force and the others, and this possibility was explored before collecting the data sets which will be analysed here. In an attempt to reduce the problems outlined above three possible heater operating strategies were considered:

- operating the test rooms in pairs each with a 12 hour ON/12 hour OFF schedule, but in antiphase. One room is heated from 6:00 to 18:00, and the other from 18:00 to 6:00 the following day. Taken individually the data sets will suffer from exactly the same problems as those collected previously, but analysing the differences in performance may allow the effects due to heater operation to be separated from those of climate. This technique has proved useful in previous work [Martin and Watson, 1990].
- operating the test rooms to a thermostatic schedule which is not a multiple of 24 hours. With this strategy the operation of the heater drifts in and out of phase with the meteorological variables, allowing the different sources of error to be separated. Again, the technique has proved useful in previous test room work [Martin and Watson, 1990].
- randomising the operation of the heater to ensure that its output does not correlate with external variables. This is a well known system identification technique which has been used in the past to identify the response of test buildings [Letherman, Palin and Park, 1982]. At the time of these experiments, however, it had not previously been used in the context of empirical model validation.

It seems that any of these operating strategies could provide data sets which would allow the source of simulation errors to be isolated more effectively. However, the first suffers from a number of disadvantages: it requires that the test rooms are configured in identical pairs for each experiment, implicitly assuming that these pairs are perfectly matched. The resulting data then has to be interpreted in terms of the difference in performance between the pair of identical buildings operated with different heating strategies. Any attempt to interpret the data in the way in which the model will normally be used, to make predictions about a single building, causes the problems identified in the earlier data sets to reappear. The second strategy appears to solve

these problems, providing data from a single test room which will, over time, allow the influence of the auxiliary heat source to be separated from that of climate. The problem with this operating strategy is that although the chosen heater operation will not be correlated with meteorological variables it still has a statistical structure of its own, in that it consists of the same sequence repeated many times, which in turn makes the extraction of the causes of simulation errors less robust.

The third strategy avoids these problems. The power of randomising input sequences for model identification has already been demonstrated numerically in Chapters 4 and 5. There it was shown that when a pseudo-random sequence was used as input the response of the system to that input could be identified with a much greater precision than when the input was, for example, a climate variable. This represents a powerful argument for the use of such sequences when designing experiments for use with the new error identification technique.

One criticism which can be levelled at such randomised operation of the room heat source is that it is unrealistic. However, the purpose of model validation is to establish that the representation of a building within the model is satisfactory over a wide range of operating conditions. In fact, the type of randomised heater operation proposed represents a very stringent test of any model. Consider a heater sequence in which the heater is either on or off, and its state at each time step is decided randomly, for example by tossing a coin. In this sequence information about the past operation of the heater provides no information about the likely future operation. This property makes the sequence a very difficult one for a dynamic simulation to follow: it minimises the chance of errors from different sources cancelling out.

One potential problem with the random sequence described is that it gives relatively little weight to the longer time constants present in buildings. Simple calculation indicates that if such a sequence is produced at hourly timesteps there will be very few periods of uninterrupted heating longer than about six hours. Since inverting a sequence of this type yields another sequence of the same type there are also few uninterrupted cooling periods longer than about six hours.

The EMC test rooms have a relatively short time constant, of between 8 and 12 hours. For these rooms the heater sequence described is appropriate. For buildings with longer time constants, however, the sequence is unlikely to stress sufficiently the slower heat transfer processes. One solution to this problem, which has now been used in a component testing context in the PASSYS project, is the Randomly Ordered Logarithmic Binary Sequence (ROLBS) [PASSYS, 1990], in which the required number of occurrences of each length of heating pulse is specified and the pulses are then arranged in random order. Whilst this sequence retains its lack of correlation with the meteorological variables driving the room, it accentuates the lower frequencies which inevitably causes unwanted auto-correlations within the test sequence. An alternative solution [Martin and Watson, 1993] uses a digital filter to emphasise the low frequencies in a white noise test sequence. Because the filter employed is known, its inverse can be used to 're-whiten' the resulting experimental data before it is analysed. However, because of the fast response of the EMC test rooms neither of these techniques was considered necessary here, and it was decided that the rooms would be operated with a straightforward randomised heater schedule of the type initially described.

6.3 Operating a simulation model for model validation

The next issue to be addressed is how the simulation model is operated to generate predictions to compare with the measured data. Early comparisons with data from rooms heated to a given setpoint for part of the day were carried out by supplying to the model the heating setpoint schedule that had been implemented in reality. This is the way the model would be used in, say, a design context, but it proved not to be a good way to operate it for validation purposes. During periods when the heating was off the heat input was, of course, correctly predicted as zero, but there were errors in the predicted temperatures as the building reacts to other influences. When the heating cycle began there were errors in the prediction of the amount of energy required to bring the room to the setpoint, and simultaneously errors in the predicted temperatures as the room warmed up. Finally, when the room and simulation model reached the setpoint there were no further errors in the predicted temperature, but there were errors in the prediction of the energy required to maintain that setpoint. It turns out that this problem

is best solved by modifying the way in which the simulation model is operated.

The difficulties in the interpretation of early validation experiments outlined centred on the way that a deterministic heating schedule was supplied to the model. For this experiment randomised operation of the test room heat source has been chosen, on the grounds that by reducing unwanted correlations between inputs it would yield data most likely to allow the errors due to different test room driving forces to be isolated. The choice of a randomised heat input points the way to a strategy for avoiding the problems of interpretation which were described. The obvious way in which to carry out a simulation to compare with this data is to apply the same pseudo-random sequence to the model of the test room. The resulting simulations of zone and surface temperatures can then be compared with the measured values to assess the quality of the simulation. We will refer to this mode of model operation as 'power scheduled'.

The heater power scheduled approach to model operation has a further important benefit when used with models such as SERI-RES in which a single combined network is used to represent the radiative and convective processes within a zone. The resulting zone temperature can only be interpreted directly as an air temperature under a fairly restricted set of operating circumstances [Haves, 1989]. When these requirements are not fulfilled, employing the simulated zone temperature as a thermostat temperature results in a simulation which is physically incorrect. In a power scheduled run these discrepancies will still be apparent when predicted zone and measured air temperatures are compared, but predicted and measured surface temperatures and heat fluxes should still be predicted correctly by the model, and can be compared directly. This represents a further significant advantage of the heater scheduled approach to operation of the model.

The principle disadvantage of heater scheduled operation of the model is that it provides no information about how good the model is at predicting the energy consumption of the room, the actual consumption having been fed to the model as part of the simulation process. Since the

prediction of auxiliary energy consumption is one of the principle applications of the model, its performance on this task is likely to be of interest. To obtain predictions of energy consumption, the model can be fed with hourly details of the temperature within a test room. If this value is used as the setpoint, the model will predict the amount of auxiliary energy (which may be positive or negative) that is required to maintain that temperature. This prediction can then be compared with the actual energy consumption measured. The model also produces predictions of the surface temperatures and heat fluxes which can, again, be compared with the measured values. This is referred to as 'temperature scheduling'.

To summarise, there are two ways of operating the simulation model. The first, heater power scheduling, is to apply the same heat input to the model as was applied to the test rooms, and observe the quality of the predictions of the resulting temperatures and heat fluxes. The second approach, zone temperature scheduling, is to force the model zone temperature to follow the air temperature measured in the test room, and examine the prediction of the energy required to do this, and of the resulting surface temperatures and heat fluxes. The consistent use of one of these strategies throughout a simulation run will eliminate one of the problems previously encountered in interpreting the results of comparisons between the measured data and simulations, namely that when the model is supplied with a setpoint schedule it is effectively predicting energy consumptions whilst the room is heated, and temperatures when it is not. In the light of the above discussion it becomes clear that in this situation the model was effectively being operated under power scheduling when the heater was off, and under temperature scheduling when it was on. This observation goes some way to explaining the resulting confusion when the results of such experiments are interpreted.

The availability of these twin strategies is in no way dependent on the fact that the room is being operated with a randomised heater schedule rather than under thermostatic control. Either technique can be used with data sets featuring any kind of heater control, including even unheated operation. The thermal model of the building fabric calculates the temperatures and heat flows which result from heater operation. In order to predict the heat input which will be

required to maintain a given temperature profile in a building it is necessary to introduce a further model of a feedback control system. Since the principal aim of the work described here is to validate the building fabric modelling in SERI-RES heater scheduling seems to be the most appropriate choice of strategy. The added advantage, outlined above, that it removes the problems inherent in interpreting the zone temperature makes heater scheduling the obvious simulation strategy for use here.

The notion of 'blind' simulation runs is central to the approach to model validation adopted in this work. The model user who attempts to predict the performance works in ignorance of the measured values which he or she is called upon to predict. The reason for insisting on this mode of operation is simple: it has been demonstrated [Martin, 1991] that it will almost always be possible for a modeller to adjust the input parameters of the test room model to bring its predictions into line with observations. In the case of unconscious adjustment bias will be introduced into the simulation results which may vary between users and between models. In the case of conscious adjustment, the inputs to the model can almost always be systematically altered (within physically reasonable bounds) to achieve a high level of agreement - empirical validation as a test of the model becomes useless.

6.4 Details of the dataset

The test rooms are described in detail in [Martin and Watson, 1988]. The data used here came from an experiment which had been designed to investigate the effect of heater position on the performance of passive solar buildings. It used six of the eight rooms, which were carpeted with a standard carpet/underlay combination, and were unventilated. Two of the rooms were equipped with single glazing, two with double glazing, and two with an opaque infill panel. Each was heated by an oil filled electric panel radiator, placed at the back in three of the rooms and at the front, beneath the window in the other three. Preliminary inspection of the data revealed quite different simulation error characteristics between the rooms with heaters at the front and those with the heaters at the rear. In order to test the new analysis technique data from the two double glazed rooms will be analysed. Table 6.2 reiterates the configuration of

those rooms during the experiment.

	Room number	
	0	1
Heater position	Rear	Front
Heater type	Panel radiator	
Ventilation rate	Zero	
Glazing	1.5 m ² Double	

Table 6.2: Test room configuration during trial

The dataset is of length 46 days. The data consists of measurements at 5 minute intervals and is continuous over the measurement period. In order to compare the data with the predictions of SERI-RES it was reduced to hourly average values, giving a total of just over 1100 points. The study of Section 4.5 suggests that this should be sufficient to identify the error response to all the forces driving the model. Table 6.3 gives the EMC volume number and the dataset start and finish dates.

EMC Volume number	Start date	Finish date
VOL 400	18th May 1991	2nd July 1991

Table 6.3: Details of data collection period

Table 6.4 summarises the climatic conditions during the trial in terms of the mean values of the key variables.

	External temperature (°C)	South vertical solar (W/m ²)	Wind speed (m/s)
VOL 400	11.9	192	3.9

Table 6.4: Summary of climatic conditions during data collection period

6.5 Carrying out simulations for comparison with the measured data

The simulations described here were based on SERI-RES descriptions of the EMC test rooms which had been prepared some years prior to data collection [Lomas and Bowman, 1987]. The only changes to those descriptions were the addition of carpet to each room, which was done using standard values for the thermal properties from [Chartered Institution of Building Services Engineers, 1986]. We can thus consider that the simulations were carried out 'blind', in accordance with the SERC/BRE methodology described in Chapter 2.

6.5.1 Meteorological data handling

Two modifications were made to the way in which the model treated solar radiation. The first of these related to the timing of weather data.

In its original form, SERI-RES accepted meteorological data which employed the American timing convention, that is the data recorded with, for example, time label 2 were the average values over the period 1:00 to 2:00 am. In the UK, meteorological data is traditionally recorded in a different format, the hour 2 values being averages from 1:30 to 2:30 am. Somewhere in the revision history of SERI-RES an undocumented change was made to the solar radiation processing code which attempted to allow the model to use UK weather files. The change and its consequences (which include the fact that the schedules and outputs sections no longer operate as described in the program documentation) have since been described in full elsewhere [Lomas, Parand and Eppel 1989].

At the EMC test site meteorological data is recorded at five minutely intervals, and from these records files conforming to either American or UK timing conventions can be prepared. It was decided that files to the original American specification would be used, and the undocumented change to the code was therefore removed, restoring the model to its original form.

The second modification concerned the way in which the model determined the amount of solar radiation incident on the glazing of the test room. SERI-RES takes as input from its

meteorological data file measurements of horizontal diffuse and direct normal solar radiation. From this information and details of the orientation of the external surfaces of a building the model calculates the direct and diffuse radiation incident on each surface.

Implicit in this process is a model of the distribution of the diffuse solar radiation component about the sky vault. SERI-RES is relatively sophisticated in this respect, and assumes a concentration of diffuse radiation around the solar disk, or circumsolar component, surrounded by a uniform diffuse background. On extremely dull days the circumsolar component is set to zero, and all diffuse radiation is assumed to come from a uniform sky [Haves and Littler, 1987]. However this is still a highly simplified representation of the real sky, and as a result errors can be expected to occur in the radiation levels predicted on the surfaces of a building. There is a further source of inaccuracy when the model is used in this way. At the EMC test site, as at many meteorological stations, direct normal radiation is not measured. Instead global horizontal and diffuse horizontal radiation are measured, and the direct normal radiation deduced from these values. However, this is a process which is, under some circumstances, highly sensitive to errors in the original measurements [Watson, 1984]. Errors in the estimated direct normal radiation in turn cause further errors in the predictions of radiation on the external surfaces of the building.

Without doubt, both of these areas are worthy of further investigation, leading to improved radiation models and more effective measurement techniques respectively. However, such investigations are clearly outside the scope of the present work, which seeks to produce the best simulations of the test room fabric possible with SERI-RES. Although the model computes the radiation incident on all of the external surfaces of the test cell, it has been demonstrated in previous sensitivity studies [Martin, 1991] that it is only the radiation landing on the glazing which has a significant influence on the performance of the room. The radiation on this plane is measured on site. Previously the measured values have been used to examine the quality of the prediction of that radiation which was being used by the model [Martin and Ruyssevelt, 1988]. The conclusion was that errors from this source alone could be significant.

For the simulations of the new data sets it was considered desirable that the possibility of such errors should be eliminated, or at least greatly reduced. There are essentially two ways in which the extra information can be introduced into the simulation. The first is to modify the code of the model to allow it to employ the measured radiation on the glazing directly in its calculations. However, any validation process which requires that changes be made to the code of a model is open to criticism, as it has not validated the model in its original form. A second way of ensuring that the model uses the correct solar radiation levels on the building glazing is to adjust the values fed to the model in their standard meteorological data file to ensure that the intermediate quantity subsequently calculated (in this case south facing vertical solar radiation) is in accordance with the values which have been measured.

Implementing this technique is essentially a problem of inverting the model radiation processor - the value of radiation on the glazing which the model is required to predict is known, and the values of direct normal and diffuse radiation which will force the model to make that prediction are required. Whilst this approach may seem rather artificial, it can be viewed simply as an alternative way of estimating direct normal and diffuse radiation from the quantities measured on site. The new method of estimating the required radiation quantities is optimal in the sense that when fed to SERI-RES the quantities happen to give the best possible estimate of solar radiation on the glazing of the test rooms. A programme was prepared to produce such estimates of direct normal and diffuse radiation from the data recorded at the EMC site. The programme accepts site meteorological data and places the derived solar radiation variables, together with external air temperature and windspeed, in a SERI-RES format weather file ready for use with the model. For the data set used here the resulting predictions of solar radiation on the south facing vertical are within 1% of the measured values, which is well within the accuracy of the sensors used to measure the quantity. The errors generally occur at low radiation levels, when the model is switching between alternative diffuse sky models.

6.5.2 Simulation scheduling

Following the discussion of the previous section the simulations were carried out as power scheduled, by feeding the power sequence applied to the rooms to the model. To achieve this it was necessary to modify the source code of the model. Given the caveats of the preceding section it was considered vital that these modifications were as non-intrusive as possible.

SERI-RES provides a facility, called the port, which allows the user to introduce specially written software routines with minimum risk of disrupting the normal functioning of the code. Essentially the port consists of a series of calls to a dummy subroutine at various stages in the course of a simulation. Using the port consists of replacing the dummy subroutine with purpose written code which carries out the desired actions. A flag is provided which allows the author of a port routine to identify the point from which calls to the code are being made.

A port routine was written which allowed the simulation to be scheduled from the actual power inputs. Table 6.5 summarises the function of this routine.

Stage in simulation	Status flag	Action of new routine
Initialisation (called only once)	1	<ul style="list-style-type: none"> • detect whether heater schedule file exists, and if so read into memory • detect whether temperature scheduling file exists and if so read into memory, • set pointers into schedules so that on future calls the programme will know which (if any) type of scheduling is in use, and what point has been reached in that schedule
Start of timestep	2	None
End of timestep	3	None
Start of each hour	4	<ul style="list-style-type: none"> • if heater power scheduling is in use copy next scheduled power into plant capacity for the zone, • if temperature scheduling is in use copy next scheduled temperature into setpoint for the zone, • increment pointer(s).
End of each hour	5	None

Table 6.5: Function of port routine for simulation scheduling

The new routine was tested in two ways:

- simulations which had been conducted previously were repeated with the scheduling facility disabled. This demonstrated that the new code was not causing unwanted side effects, and
- when, for example, heater scheduled simulations were conducted the energy delivered to the room during the simulation was compared with the actual energy input on an hour-by-hour basis. In all cases these were identical, indicating that the scheduling process was functioning correctly.

6.5.3 Modification to predicted floor surface quantities

Due to practical considerations, it was not considered feasible to measure floor surface temperature or heat flux on top of the test room carpet, and the measurements were instead made at the carpet/floorslab interface. In contrast, the surface quantities predicted by SERI-RES relate to the upper surface of the carpet. Before comparing simulated and measured quantities it is thus necessary to correct either the simulated or measured results.

Correcting the measurements is appealing, as the resulting comparison will then be between temperatures at the carpet/air interface, which would normally be regarded as the most natural definition of the floor surface. However, since the exact parameters describing the actual physical installation are not known, any such correction would, of necessity, be approximate. The exact parameterisation being used within the model is known, and its predictions can therefore be corrected exactly. Accordingly, it was decided that the model output should be adjusted to make its predictions compatible with the measurements made.

Figure 6.3 shows the situation modelled, and shows the location of the predicted and required floor surface temperature and heat flux. From the figure it is clear that the heat flux predicted at the measurement point is the same as that predicted at the carpet upper surface. This is a consequence of modelling the carpet as a layer without mass - heat flow into one side must be balanced by a corresponding flow out of the other side. Thus no correction to the predicted floor surface heat flux is required. The predicted temperature at the measurement point can be determined from the temperature drop which must occur across the carpet resistance, recalling that a positive heat flux flows out of the test room:

$$T'_{\text{floor surface}} = T_{\text{floor surface}} - F_{\text{floor surface}} R_{\text{carpet}}$$

where:

$T'_{\text{floor surface}}$ is the modified temperature, corresponding to the floorslab/carpet interface,

$T_{\text{floor surface}}$ is predicted temperature, corresponding to the carpet/air interface,

$F_{\text{floor surface}}$ is the floor surface heat flux, and

R_{carpet} is the resistance of the carpet layer.

The software used to process the output from the simulation model was modified to make this correction to the simulated values. In all the comparisons which follow the term floor surface temperature thus refers to the temperature at the floorslab/carpet interface.

6.6 Comparisons between simulation results and measured data

Figure 6.4 shows the predicted and measured zone temperatures in the double glazed test room with the heater placed at the rear. The temperature within the room is rather higher than might normally be experienced in a real building, peaking at almost 45°C. A number of factors contribute to this:

- the room has a large area of glazing, in order to stress the heat transfer processes associated with windows,
- in the interests of realism, the room heater was sized using a rule of thumb which assumes an internal temperature of 20°C and a worst case external temperature -1°C. Since the experiment was carried out between May and July, external temperatures were considerably higher than this (Table 6.4 indicates that the mean was actually 11.9°C), and finally
- the very nature of the pseudo-random heater schedule means that, on occasion, the heater will be operating at times of high insolation, and that these energy inputs will combine to produce abnormally high temperatures. Any attempt to eliminate this by not operating the heater when insolation is high would of course introduce unwanted correlations between heater operation and solar radiation.

Figure 6.5 shows the corresponding difference, or simulation error, curve and confirms that, as in previous studies in these rooms [Martin, 1991], a significant portion of the error follows operation of the heater. The figure also reveals a slight shift in the response to solar radiation, and a longer term trend in the error which may be attributable to the slowly changing external temperature.

Figures 6.6 and 6.7 show the corresponding curves for the room floor surface temperature. As expected the mass of the floorslab acts to filter out the response to the high frequency heater

operation in both the simulation and reality, reducing the magnitude of the errors observed previously. Figures 6.8 and 6.9 show the results for the floor heat flux. Once again this shows clearly the errors in response to heater operation, indicating that the output of the panel radiator is not being correctly coupled to the massive floor. Table 6.6 summarises the discrepancies between the simulated and measured results over the whole 46 day experimental period.

	Mean value	Mean error	Mean absolute error	Standard deviation of error
Zone temperature (°C)	29.1	1.6	1.8	1.6
Floor temperature (°C)	28.3	0.8	1.1	1.1
Floor heat flux (W/m ²)	7.6	1.1	3.5	4.6
Back wall temperature (°C)	28.9	1.2	1.5	1.4
Ceiling temperature (°C)	27.7	2.6	2.6	1.5

Table 6.6: Simulation errors for double glazed test room: heater at rear

Figure 6.10 shows the predicted and measured test room zone temperature in the double glazed room with the heat source placed at the front. Again Figure 6.11 shows the difference between the simulated and measured values. Agreement is generally worse than that observed with the heater at the rear of the room on Figure 6.4. This confirms earlier conclusions from rooms without carpet [Martin, 1991].

Figures 6.12 and 6.13 show the floor temperature predictions when the heat source is placed at the front of the room. Once again the errors are larger than those with the heater at the back of the room, and they appear to be different in character. This will be explored in more detail when the errors are analysed using the new technique in the next section. Figures 6.14 and 6.15 show the floor heat flux results. Once again the errors are larger than those obtained when the

heater is placed at the rear of the room. Table 6.7 summarises the results for all the simulated and measured quantities from the test room with the heat source at the front.

	Mean value	Mean error	Mean absolute error	Standard deviation of error
Zone temperature (°C)	28.9	2.0	2.1	1.7
Floor temperature (°C)	26.2	3.0	3.0	1.4
Floor heat flux (W/m ²)	6.8	2.0	4.9	5.8
Back wall temperature (°C)	27.1	3.1	3.2	1.6
Ceiling temperature (°C)	27.4	3.0	3.0	1.6

Table 6.7: Simulation errors for double glazed test room: heater at front

An important conclusion can be drawn from the results shown in Tables 6.6 and 6.7. The model over-predicts both temperatures and surface heat fluxes. However, the degree of over-prediction is less severe in the room with the heater placed at the rear, confirming the conclusions drawn from the hourly data on Figures 6.4 to 6.15. The most obvious physical explanation of this effect is in terms of the mixing of the room air which will occur when the warm air rising from the heater at the rear of the room is carried across the ceiling and descends past the test room window panel, forming a convective loop. When the heater is placed at the front of the room such mixing is likely to be greatly reduced, the upward air movement from the heater being counteracted by the cold air flowing down the window.

6.7 Detailed analysis of simulation error structure using the new method

In this section the new cross-correlation/deconvolution analysis will be applied to the simulated and measured data from the two double glazed test rooms.

The technique will be used to extract the impulse responses of the prediction errors observed within the test room to forces driving the room: heater power, external temperature, solar radiation, wind speed and sky temperature. Together with the measured values of the driving forces, those impulse responses are used to reconstitute the prediction error. In this way the relative contributions to the error can be extracted, together with the proportion of that error which is not explained by the chosen driving forces.

6.7.1 Zone temperature predictions

Figure 6.16 shows the relative contributions to the error in predicted zone temperatures in the double glazed test rooms. As in previous work the most significant contribution appears to be from the operation of the room heater, with smaller contributions from external temperature and solar radiation. The relative magnitudes of these errors are essentially independent of the position of the heater within the room. Figures 6.17 and 6.18 show the actual error responses which were obtained in the two cases, and confirm that the errors in zone temperature response to all driving forces are similar for both heater positions. The figures also indicate that the errors due to heater operation occur immediately the heater is operated.

6.7.2 Surface temperature predictions

Figure 6.19 shows the corresponding results for the errors observed when floor temperature is predicted. It is seen that the overall error in floor temperature prediction is smaller when the heat source is at the rear of the room, and it is also clear that this is principally due to the reduction in error due to heat input when the heater is at the rear. As discussed previously, there is a simple physical interpretation of this behaviour in terms of the greatly improved mixing due to a convective loop when the heater is at the rear of the room. Examining the error responses which generated this figure on Figures 6.20 and 6.21 shows that, in contrast to the results obtained for zone temperature, the nature of the floor temperature error response is strongly dependent on heater position. As expected the error is much more profound when the heater is placed at the front of the room. Over the first few hours the responses are completely different due to the different form of the heater interactions with the surfaces of the rooms. At

longer time delays the two curves adopt the same form, suggesting that there may be a small error in the model representation of the conductive effects in the room floor, an error which is, of course, independent of heater position.

6.7.3 Surface heat flux predictions

Figure 6.22 shows the breakdown of errors in floor heat flux prediction. Once again, the dominant source of error with the heater at the front of the room is the operation of that heater. When the heater is moved to the back this error again diminishes. This serves to confirm the conclusions obtained from the temperature data. However, there is also a significant contribution to the error in floor temperature from solar radiation input. The magnitude of this error changes slightly as the heater is moved, which may be due to 'optical' effects. The heater is large in comparison to these rather unrepresentatively small rooms, and moving it around clearly affects the amount of incoming shortwave radiation which reaches the floor. This may indicate that, for example, furniture could be expected to have a critical effect on the performance of floor mass. Figures 6.23 and 6.24 show the responses used to construct Figure 6.22. As expected from the temperature results the figures show a floor heat flux error which is again strongly related to heater position.

6.8 Analysis of predicted and measured responses to heater power

The analysis presented so far has concentrated on deriving the impulse responses of simulation errors - indeed this is the central concept which underlies the technique presented in this thesis. The analysis which has been developed, however, can equally well be used to extract the impulse responses of simulated and measured quantities themselves. In Chapter 2 it was indicated that comparing such impulse responses represented a powerful validation technique. Whilst this is strictly outside the scope of the work presented here, which centres on modelling simulation *error*, it will now be briefly explored.

In the previous section it was demonstrated that the most significant simulation errors occur when the test room heat source is operated. This observation can be quantified by comparing

the impulse responses of the SERI-RES predictions with those derived from the data measured in the room.

Figure 6.25 shows the response of the simulated zone temperature to heater operation, and also shows the responses derived from data measured with the heat source placed at the rear and front of the test room. The figure shows clearly the effects which were detected in the previous section: the model demonstrates a much more rapid response to heater energy than either of the two real rooms. Furthermore, the real rooms demonstrate that the response of zone temperature to heater input is largely independent of heater position, implying that it will be possible to generate a correction for it which is also independent of heater position.

Figure 6.26 shows the corresponding results for floor surface temperature. The simulated response shows a reasonable agreement with the response measured, although there is a slight difference in phase. When the heater is at the front of the room, however, there is clearly a large discrepancy between simulated and measured results. Finally, Figure 6.27 shows the results for floor heat flux. As expected, it confirms the conclusions of the previous figure, showing that floor heat flux is modelled approximately correctly when the heater is at the rear of the room, but that there are significant errors when the heater is below the window.

6.9 Conclusions of the empirical validation experiment

An empirical validation experiment has been designed and carried out in a pair of outdoor test rooms. It has been argued that such buildings provide a useful degree of realism (they are of realistic size and construction and are exposed to real climate) but at the same time they limit the complexity, and hence uncertainty, associated with an experiment (they are single zone and unoccupied). The experiment conducted here featured a pair of such test rooms. These were double glazed, and were heated by panel radiators. In one room the radiator was placed beneath the window, and in the other it was on the opposite wall.

The analytical test of the new analysis technique presented in Chapter 4 revealed that the tool

was most powerful when identifying the response to randomised inputs. In the case of an outdoor test facility the climatic variables are beyond the control of the experimenter, but other inputs which determine the performance of the room can be chosen to maximise the value of the experiment. The most obvious of these is heater operation. For the experiment described the heaters were operated to a pseudo-random schedule.

The performance of the test rooms was predicted using the simulation program SERI-RES. The simulations were carried out 'blind' by using a description of the test building which had been prepared prior to the data collection, ensuring that there was no possibility of the representation of the building within the model being influenced by a knowledge of the required results. The simulations were conducted by supplying to the model, on an hour by hour basis, the power input to the rooms which had been measured during the experiment. The model was used to predict the test room zone temperature, surface temperatures and surface heat fluxes for comparison with the corresponding measured data.

Preliminary comparisons between simulation results and measured data revealed that the model tended to overpredict both temperatures and heat fluxes in both of the test rooms. The degree of overprediction was less severe in the room with the heater at the back than when the heater was located beneath the window. It has been suggested that this may be due to the fact that in the former case a convective loop forms between the heater and window which promotes better mixing of the air within the room, bringing its performance more closely into line with the uniform air temperature assumption in the model.

The new technique was then applied to the simulation errors for both rooms, and the relative contributions to the error from the various forces driving the performance of the rooms determined. In both cases it was found that the linear time-invariant error model could explain the bulk of the observed error. It was seen that for the room with the heater at the front (beneath the window) the predominant source of air temperature prediction error was the operation of the room heater. When the heater was placed at the rear, however, the contribution

from this source was much smaller, resulting in the smaller overall prediction errors noted previously.

For both rooms the air temperature prediction error response was instantaneous: an error appeared as soon as the heater was operated. When other quantities were examined the technique proved capable of revealing subtle dynamic differences in the error structures for the two rooms. The error in predicted floor temperature is consistently positive (implying overprediction) when the heat source is located at the front of the room. With the heat source at the back it is initially zero, becomes negative one hour after the heater is operated and then becomes positive. At this point it decays in the same way as for the room with the heater at the front. When floor heat flux is examined a similar story emerges. With the heat source at the front of the room the initial overprediction decays over the first few hours to produce subsequent underprediction. With the heater at the rear of the room, however, the initial error is negative, followed by overprediction which decays over the following few hours. As with floor temperature, the overall error is smaller than when the heater is placed at the front.

It is clear from the above discussion that the highly simplified treatment of the room air and heat source within SERI-RES can be the source of significant prediction errors, in particular when the heater is positioned below the room window, a common choice of heater location. Other simulation models which are currently in use employ a similar representation. A better model of how such a heater interacts with a room could therefore produce a significant improvement in the quality of the predictions of these models. In particular, such a model will need to take into account the position of the heater in relation to the room glazing.

FIGURES

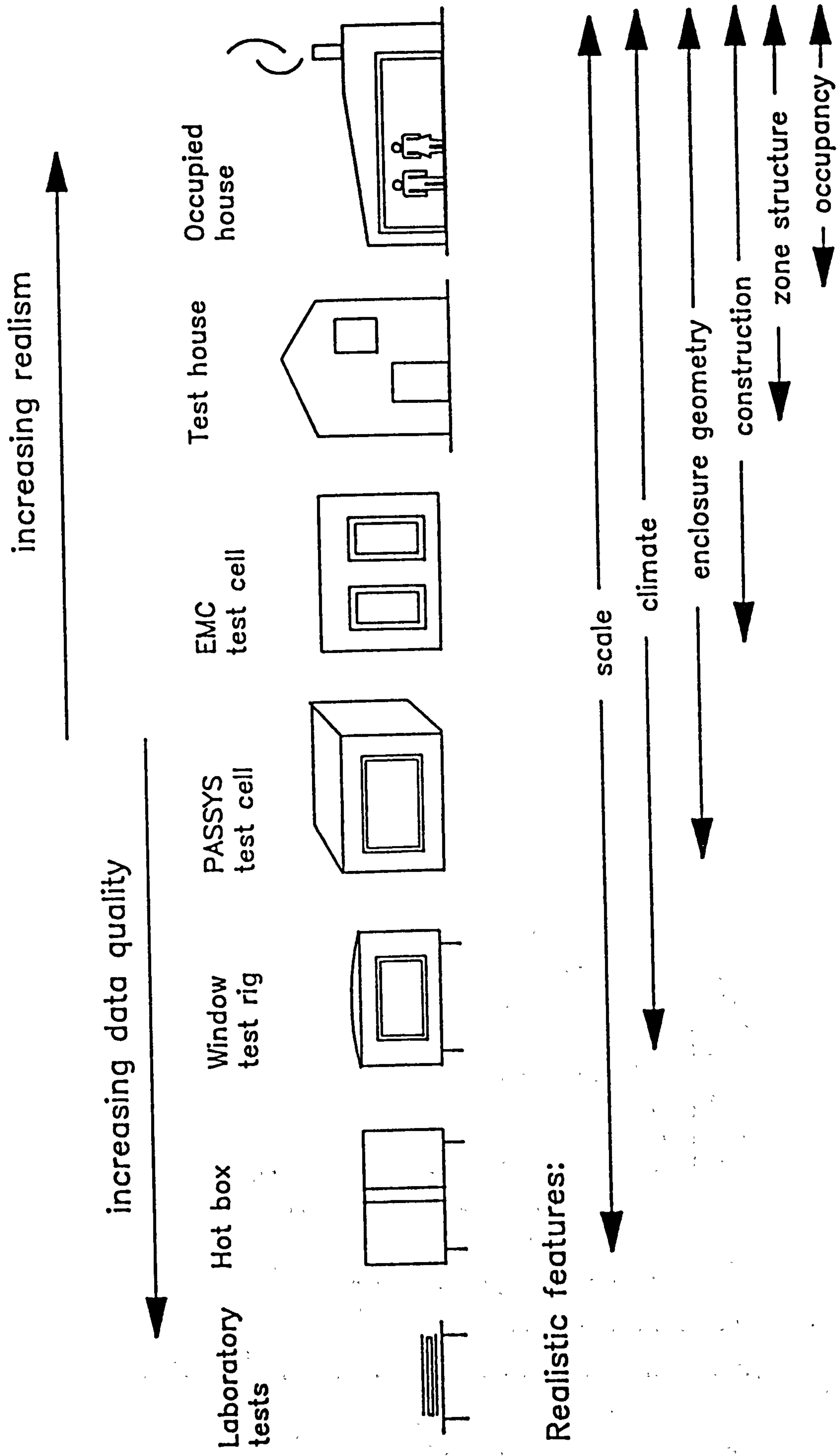


Figure 6.1: Classification of test facilities by level of realism

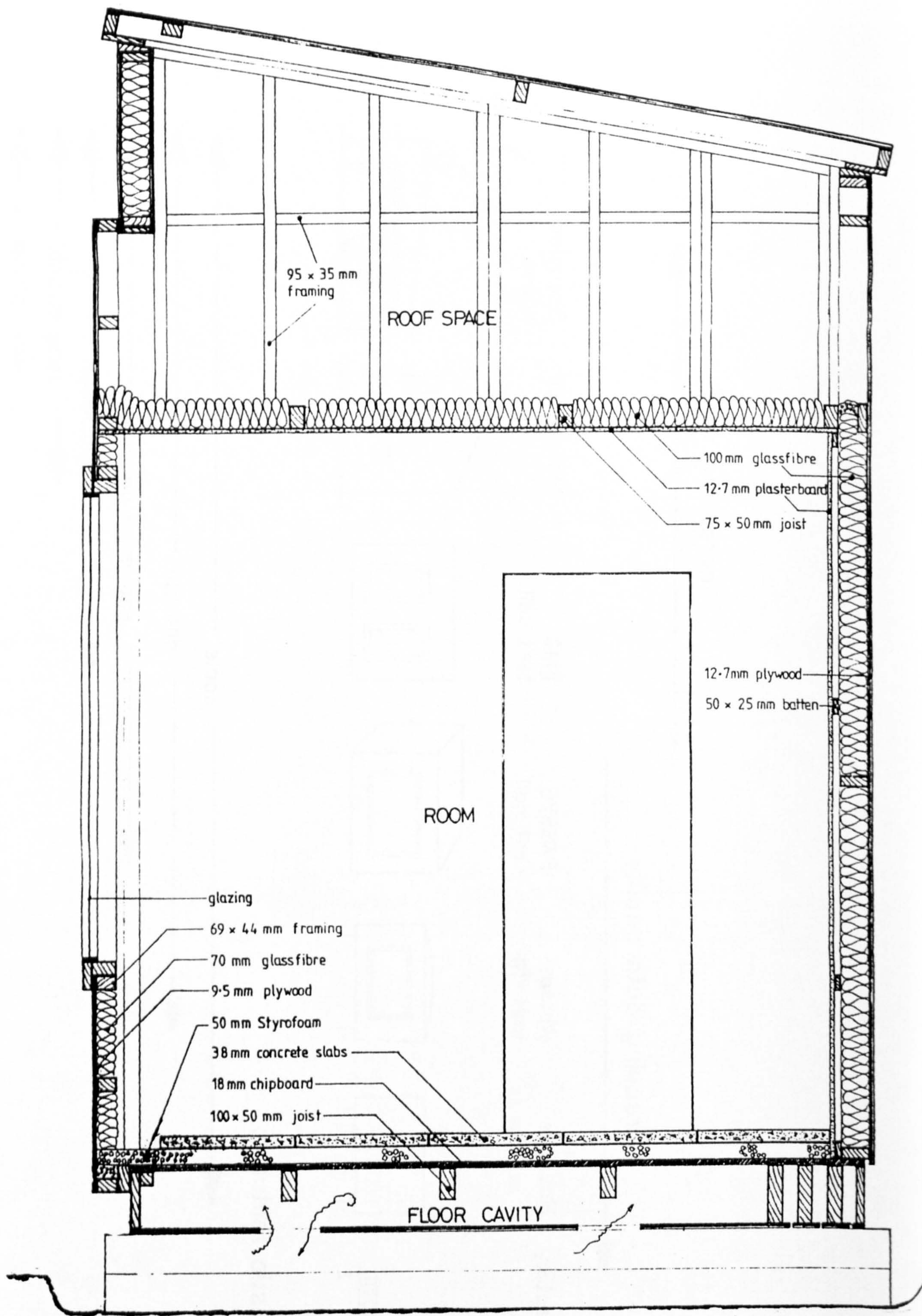


Figure 6.2: Section through an EMC test room

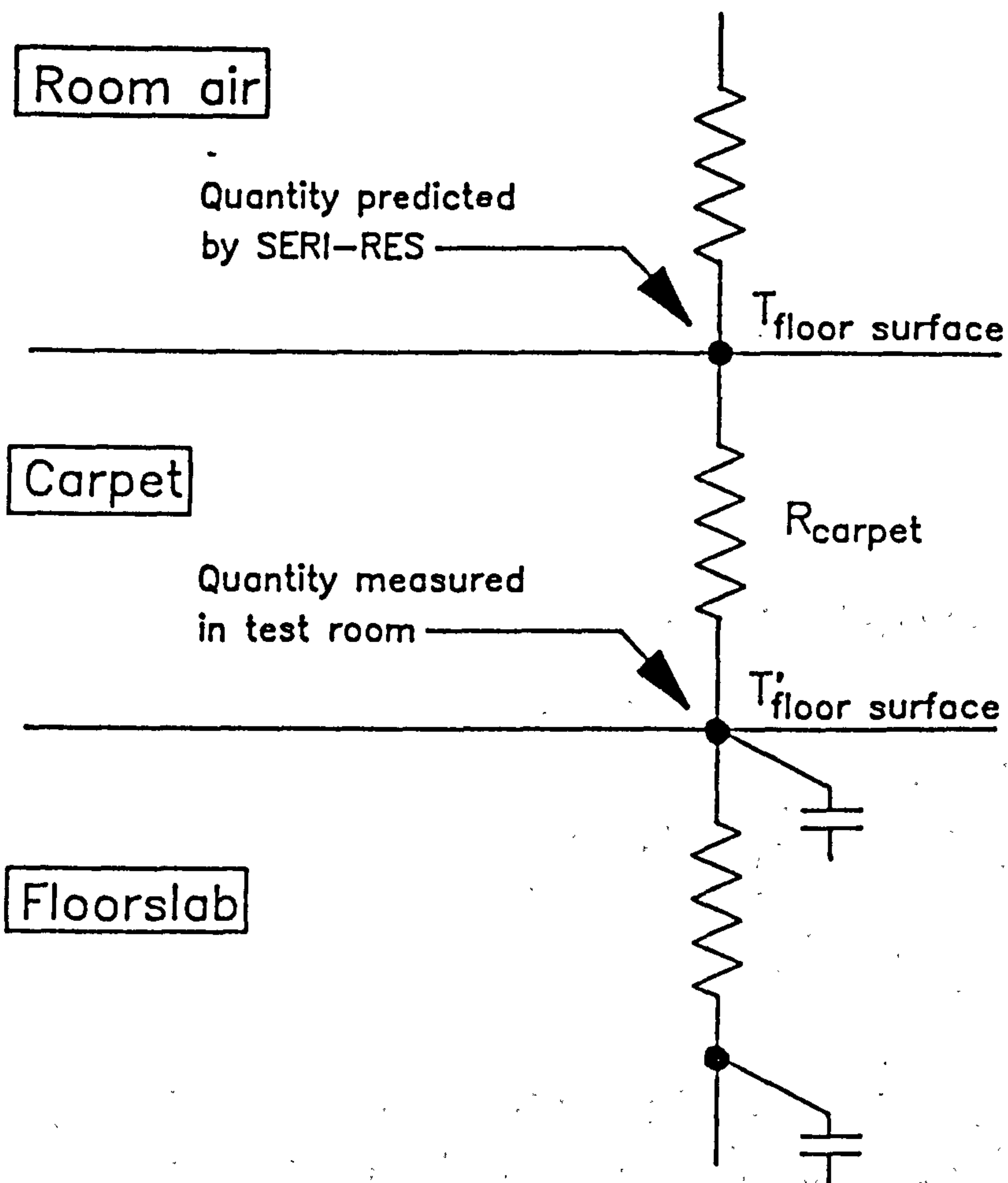


Figure 6.3: Adjustment of predicted floor surface temperature and heat flux

Predicted and measured zone temperature

Double glazed room with heat source at rear

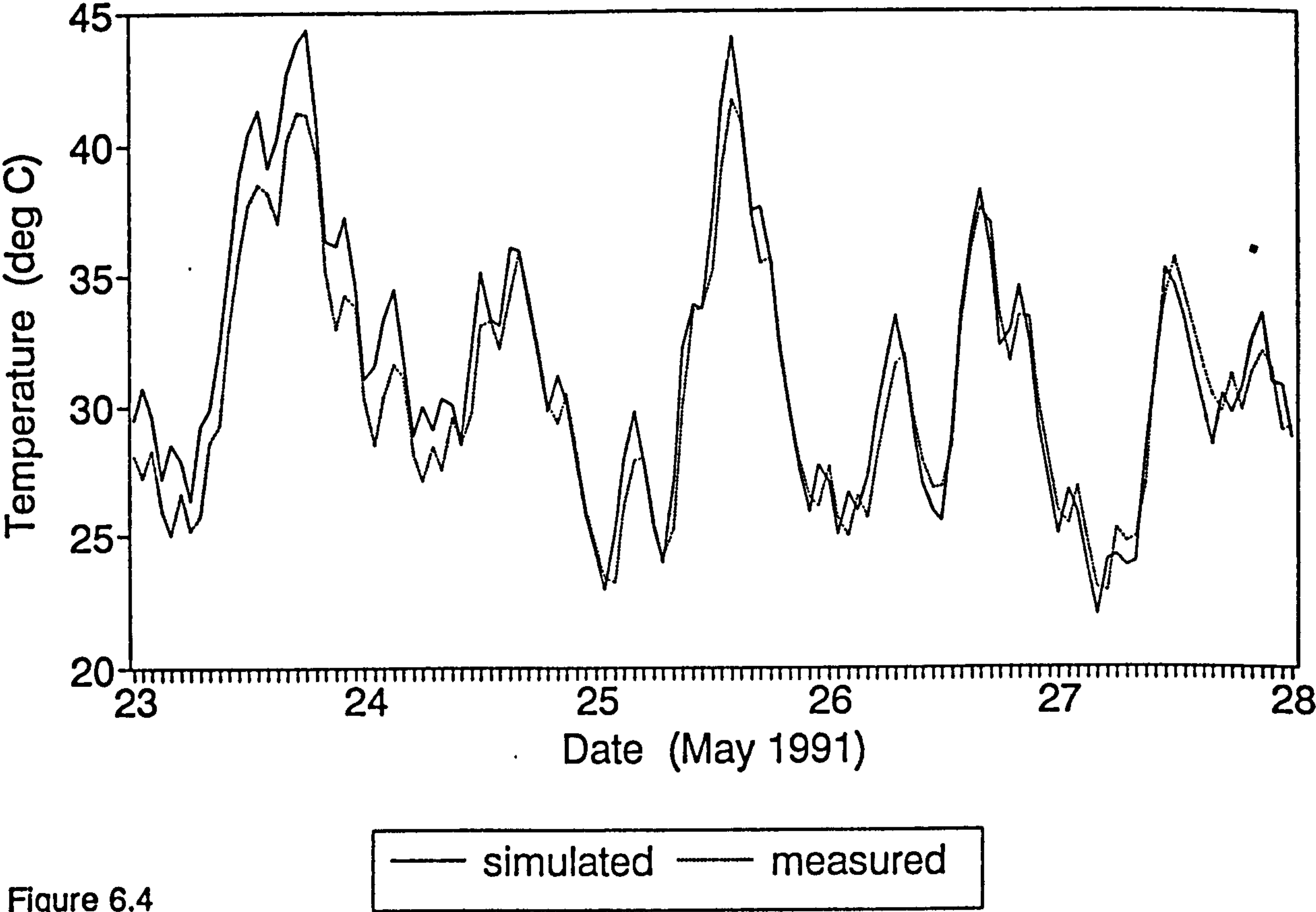


Figure 6.4

Zone temperature prediction error

Double glazed room with heat source at rear

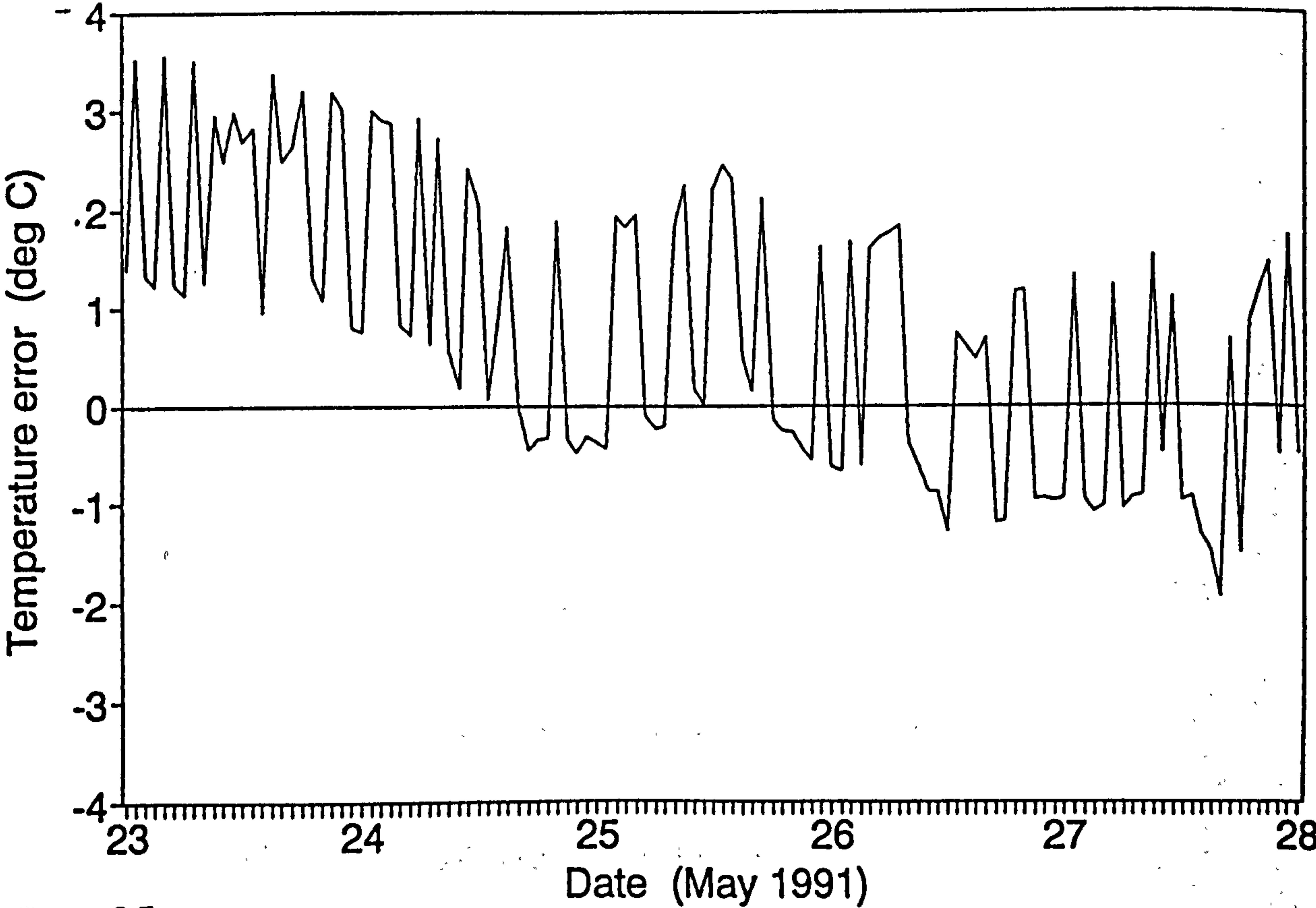


Figure 6.5

Predicted and measured floor temperature

Double glazed room with heat source at rear

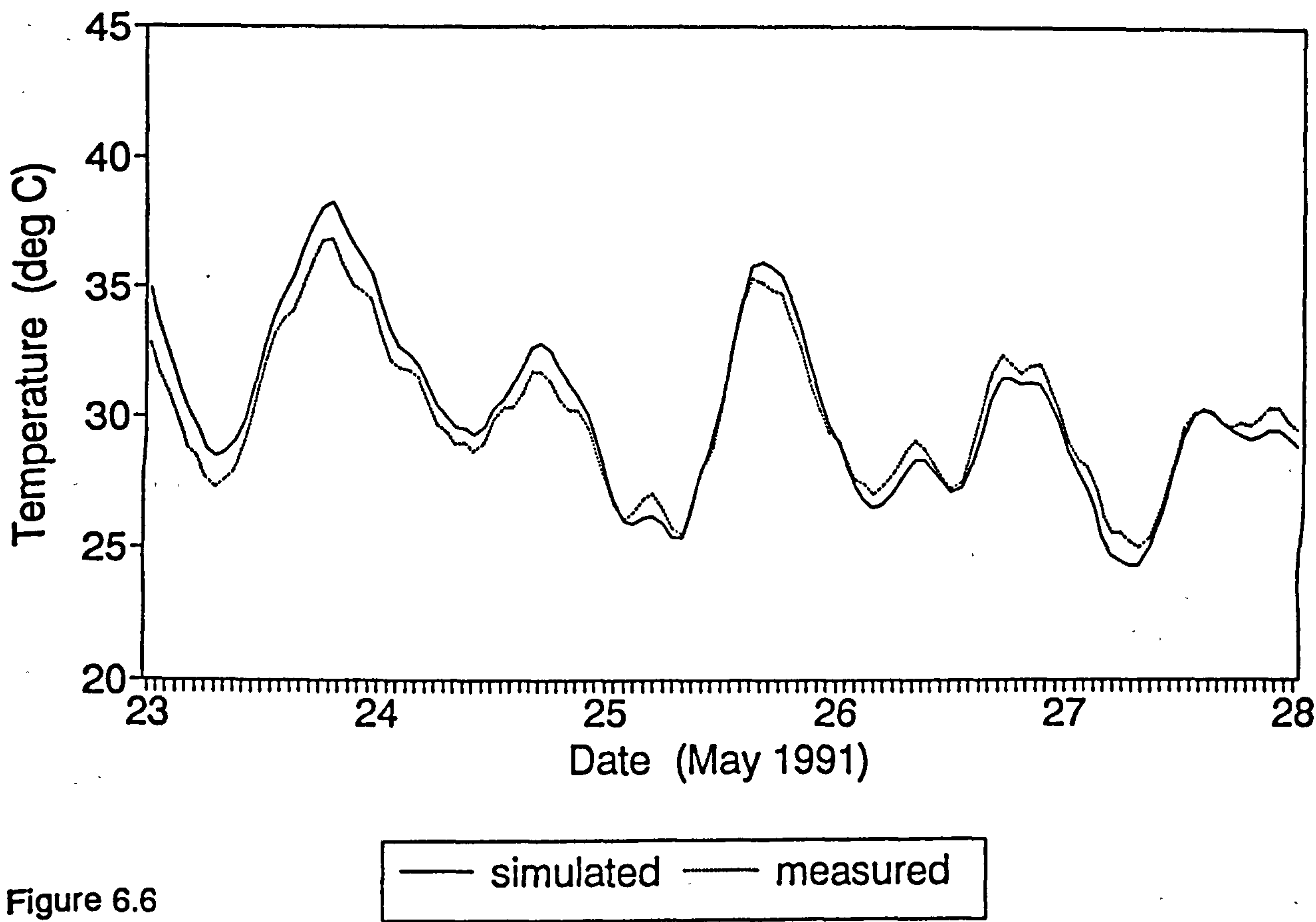


Figure 6.6

Floor temperature prediction error

Double glazed room with heat source at rear

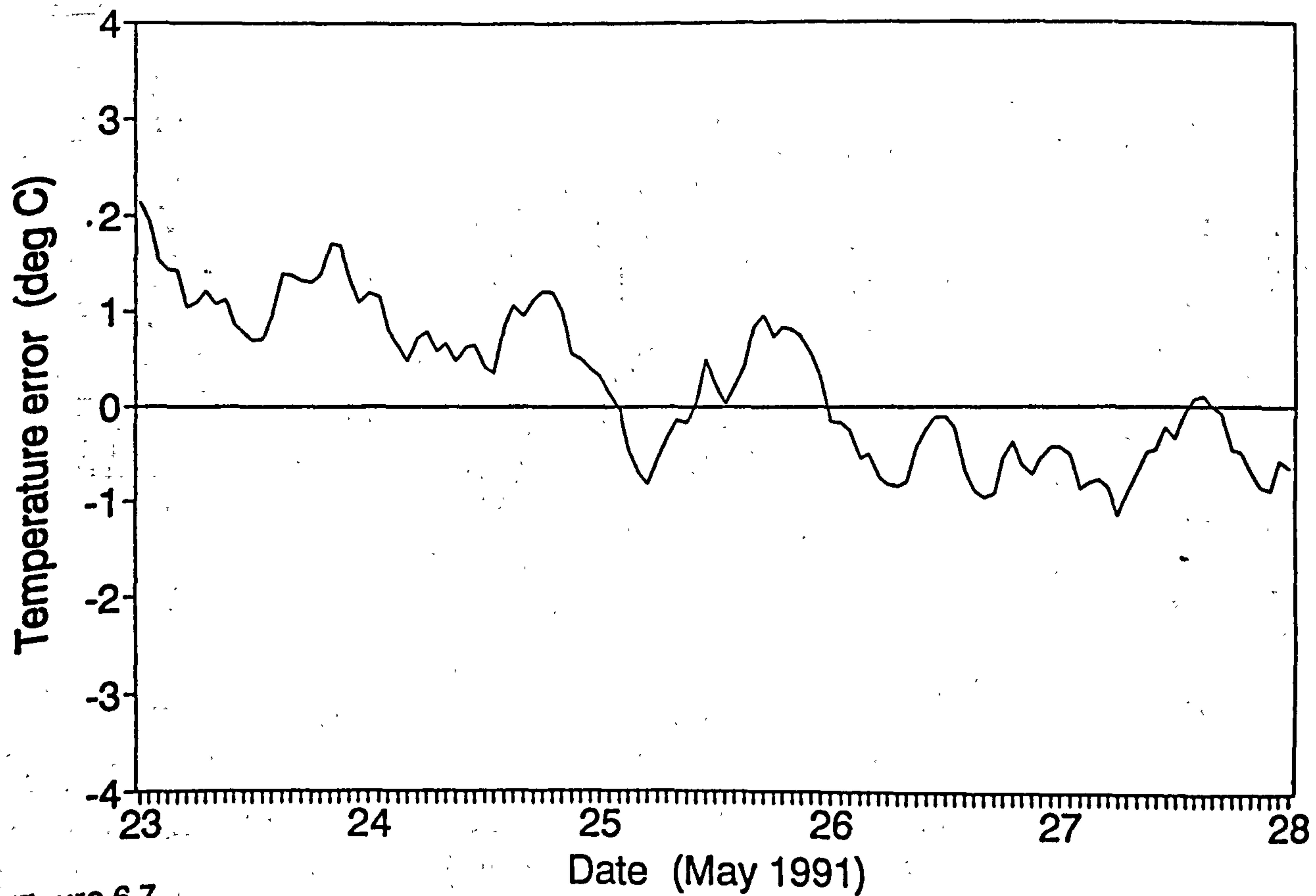


Figure 6.7

Predicted and measured floor heat flux

Double glazed room with heat source at rear

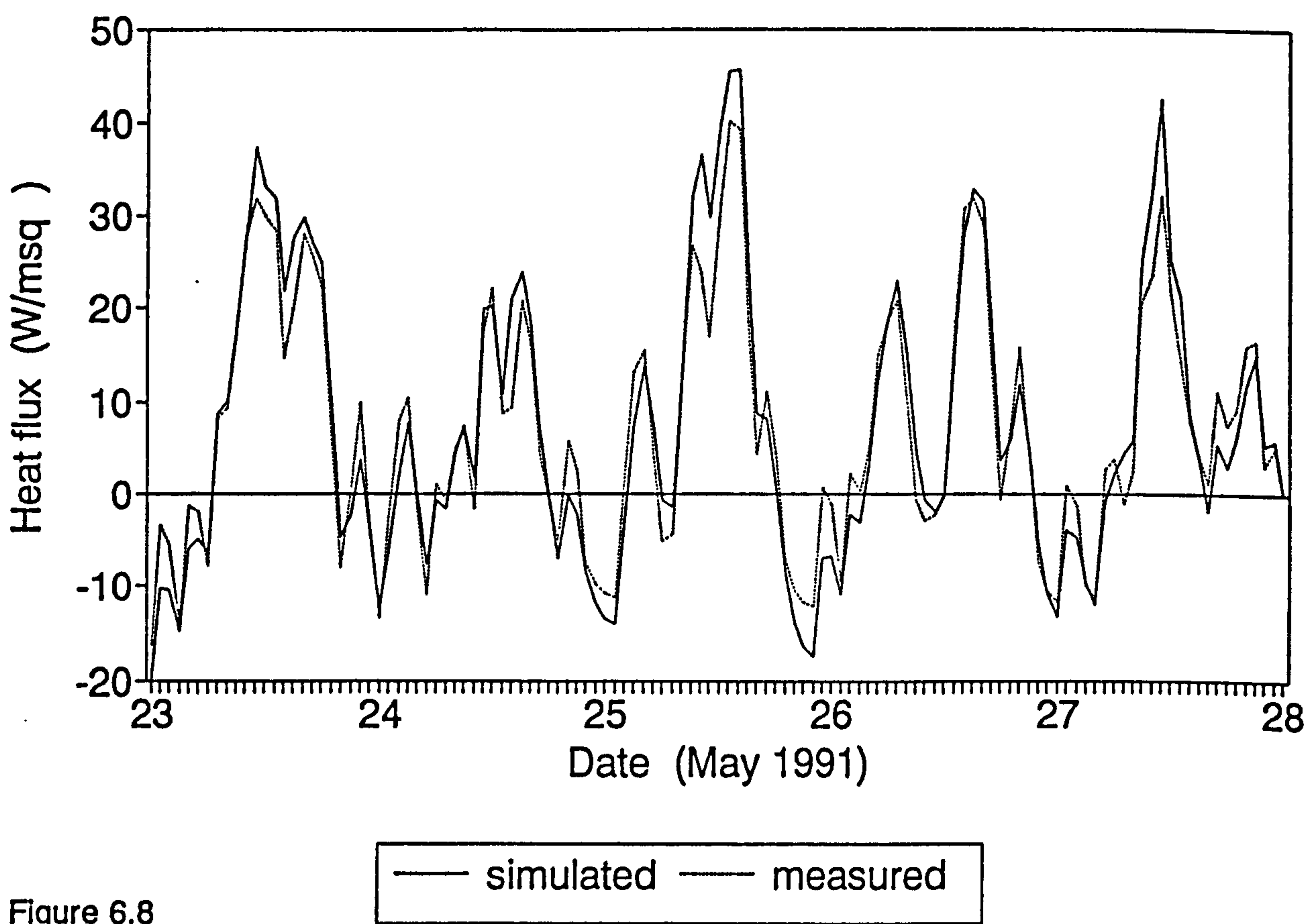


Figure 6.8

Floor heat flux prediction error

Double glazed room with heat source at rear

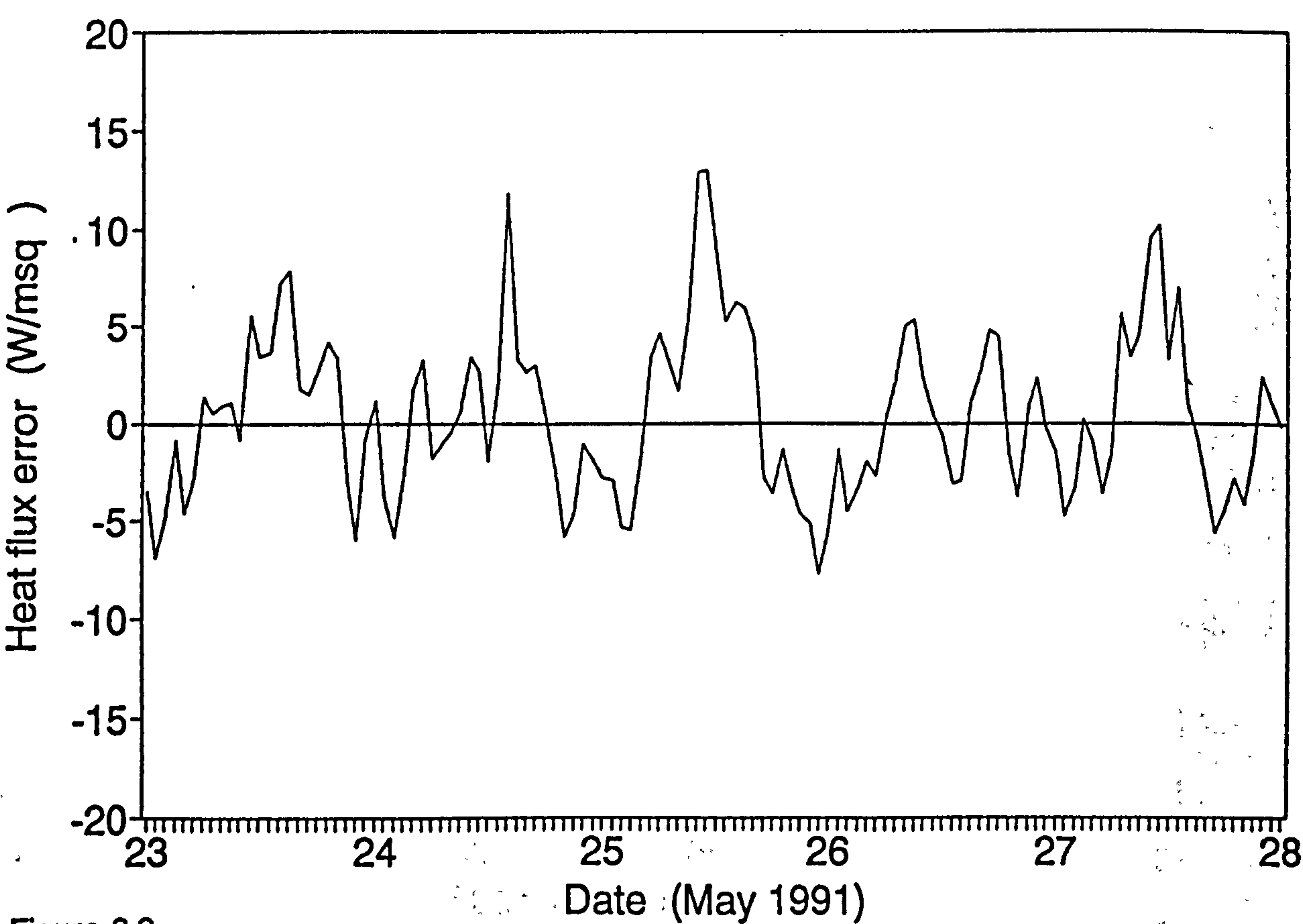


Figure 6.9

Predicted and measured zone temperature

Double glazed room with heat source at front

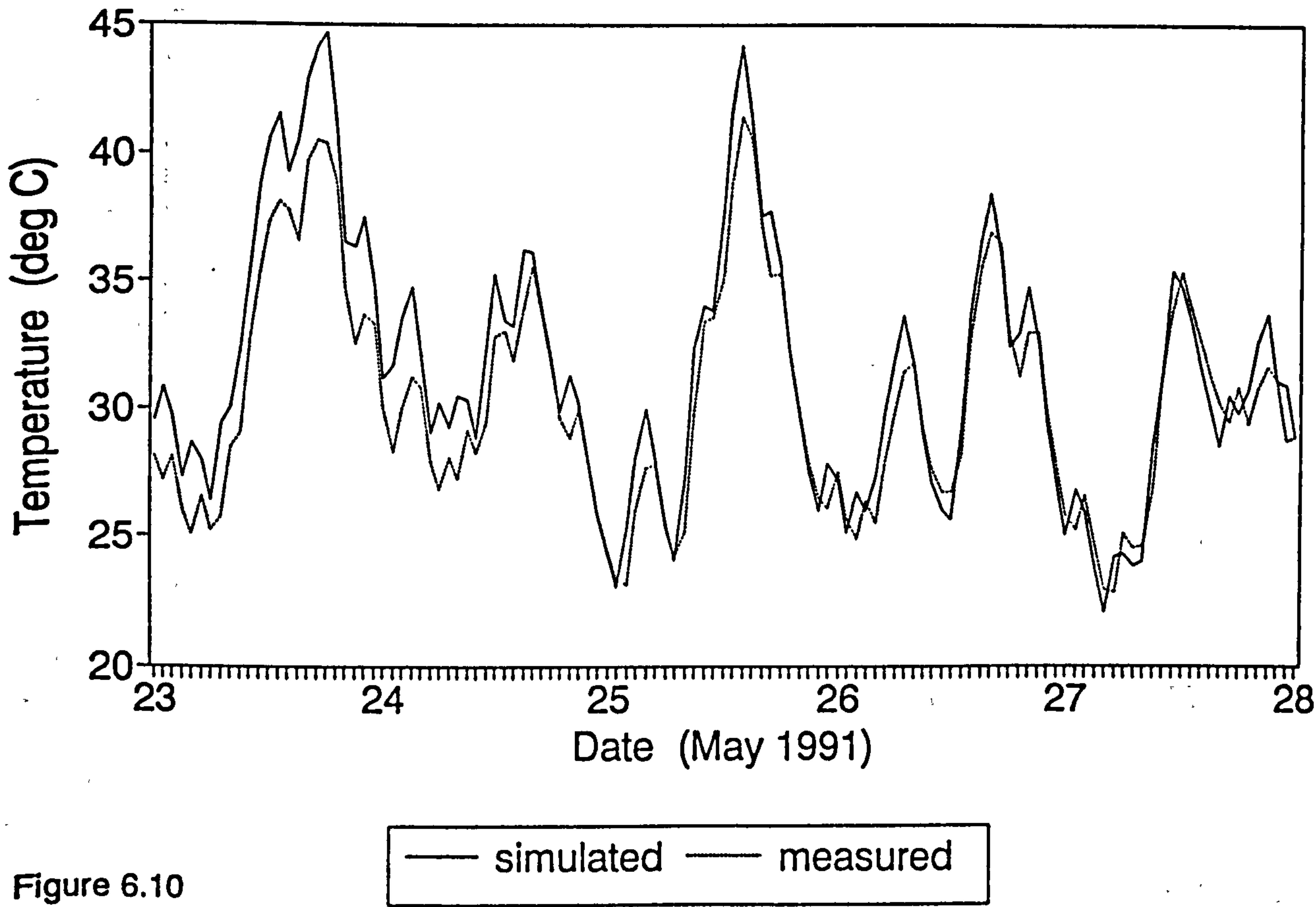


Figure 6.10

Zone temperature prediction error

Double glazed room with heat source at front

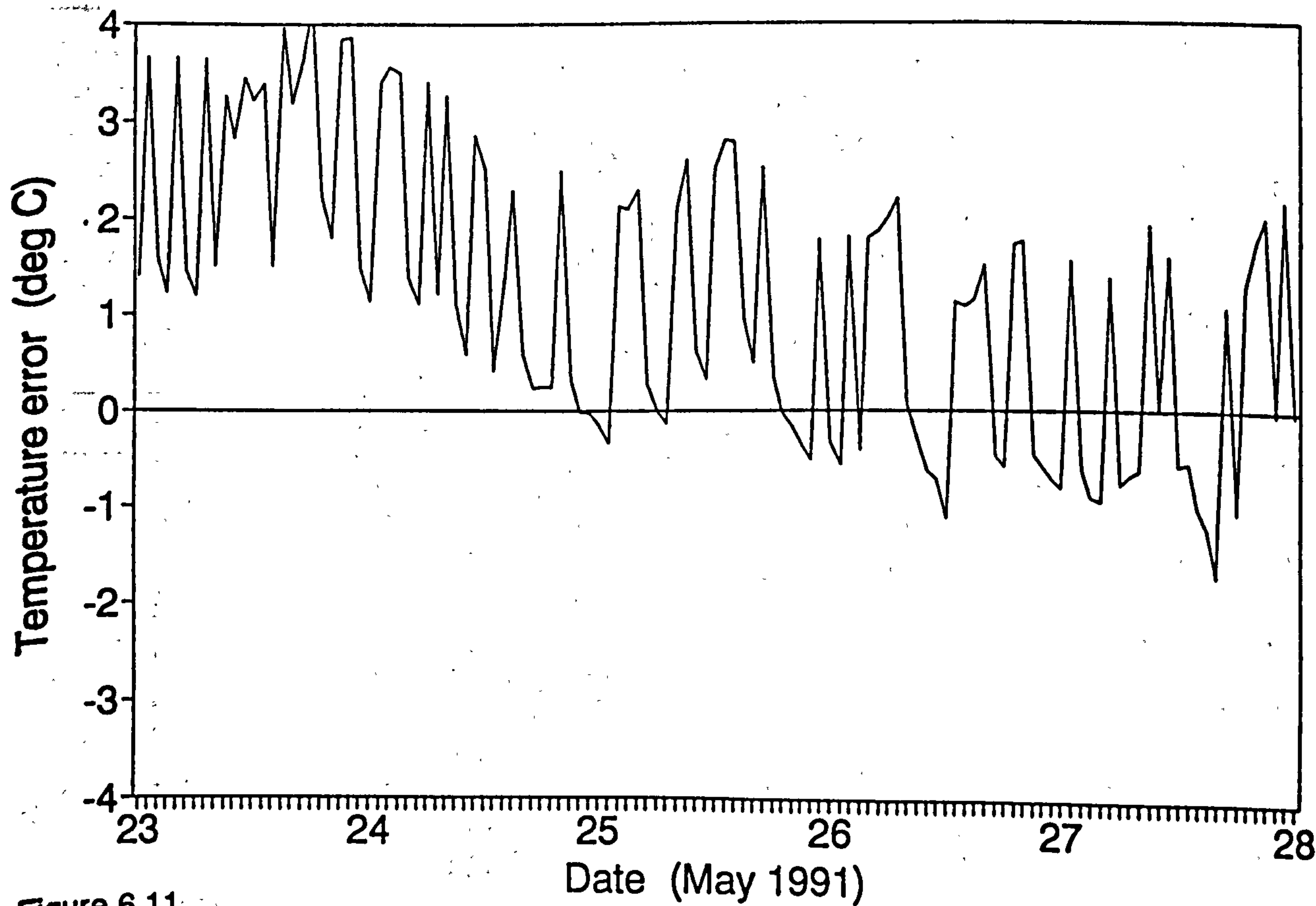
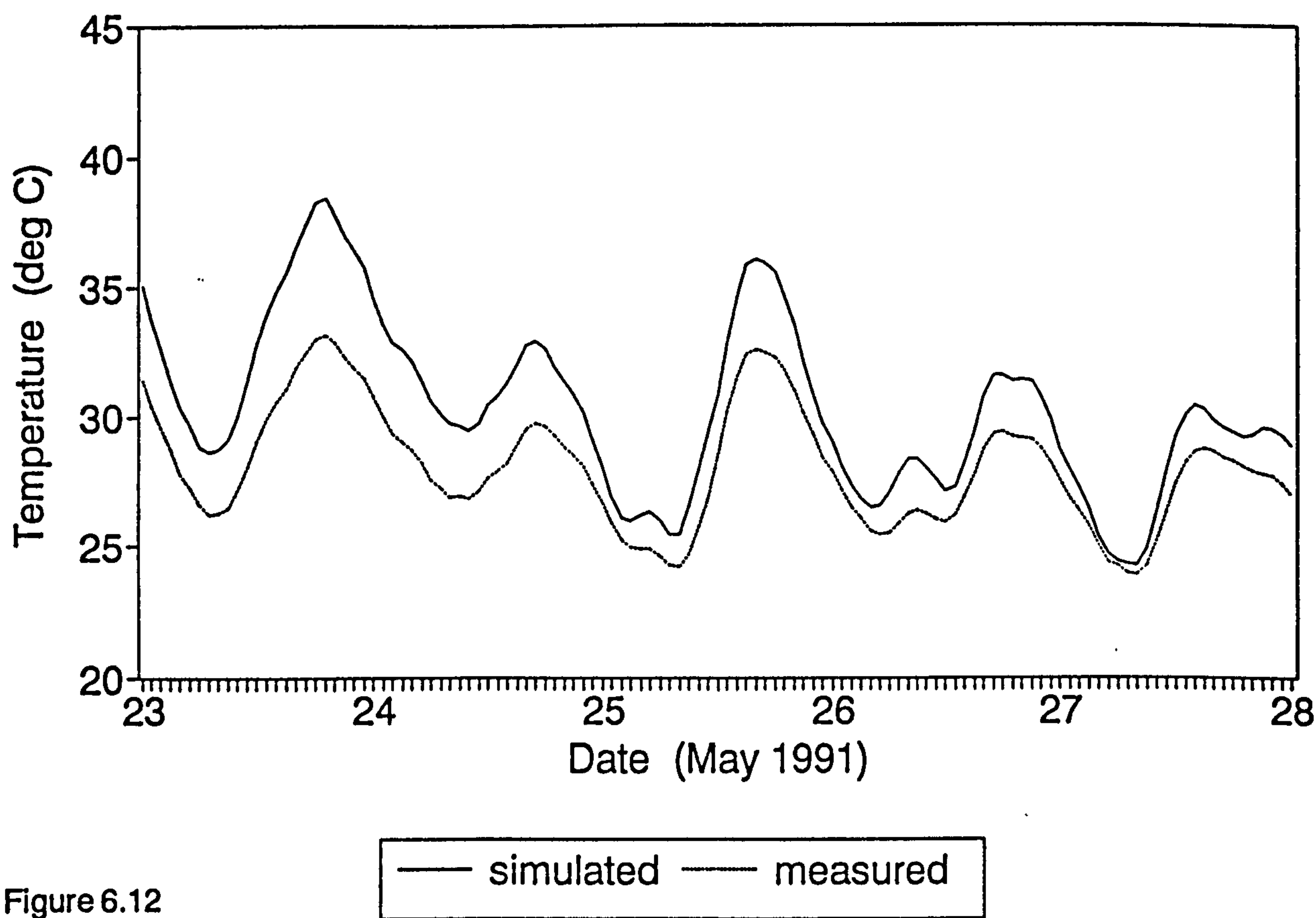


Figure 6.11

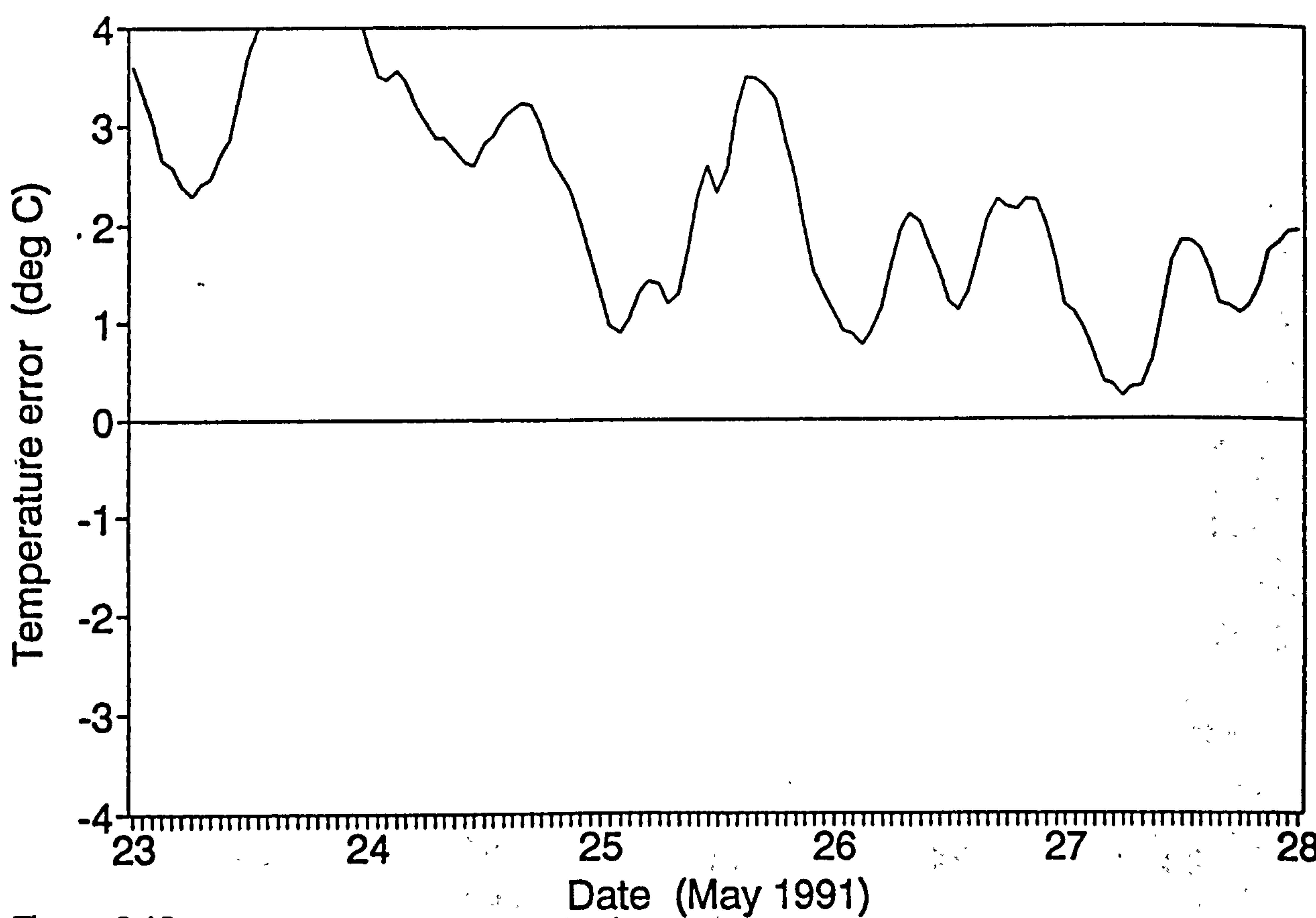
Predicted and measured floor temperature

Double glazed room with heat source at front



Floor temperature prediction error

Double glazed room with heat source at front



Predicted and measured floor heat flux

Double glazed room with heat source at front

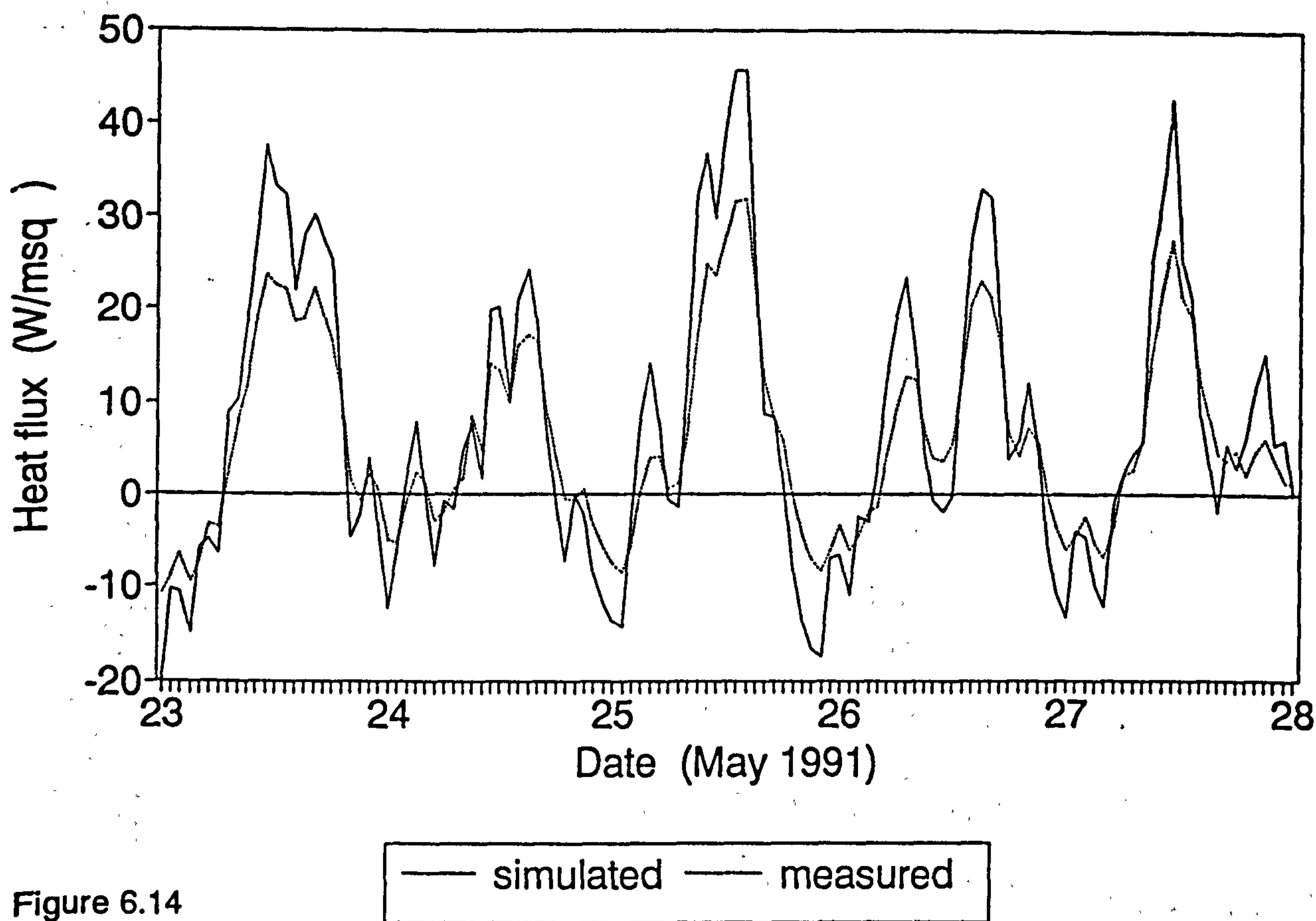


Figure 6.14

Floor heat flux prediction error

Double glazed room with heat source at front

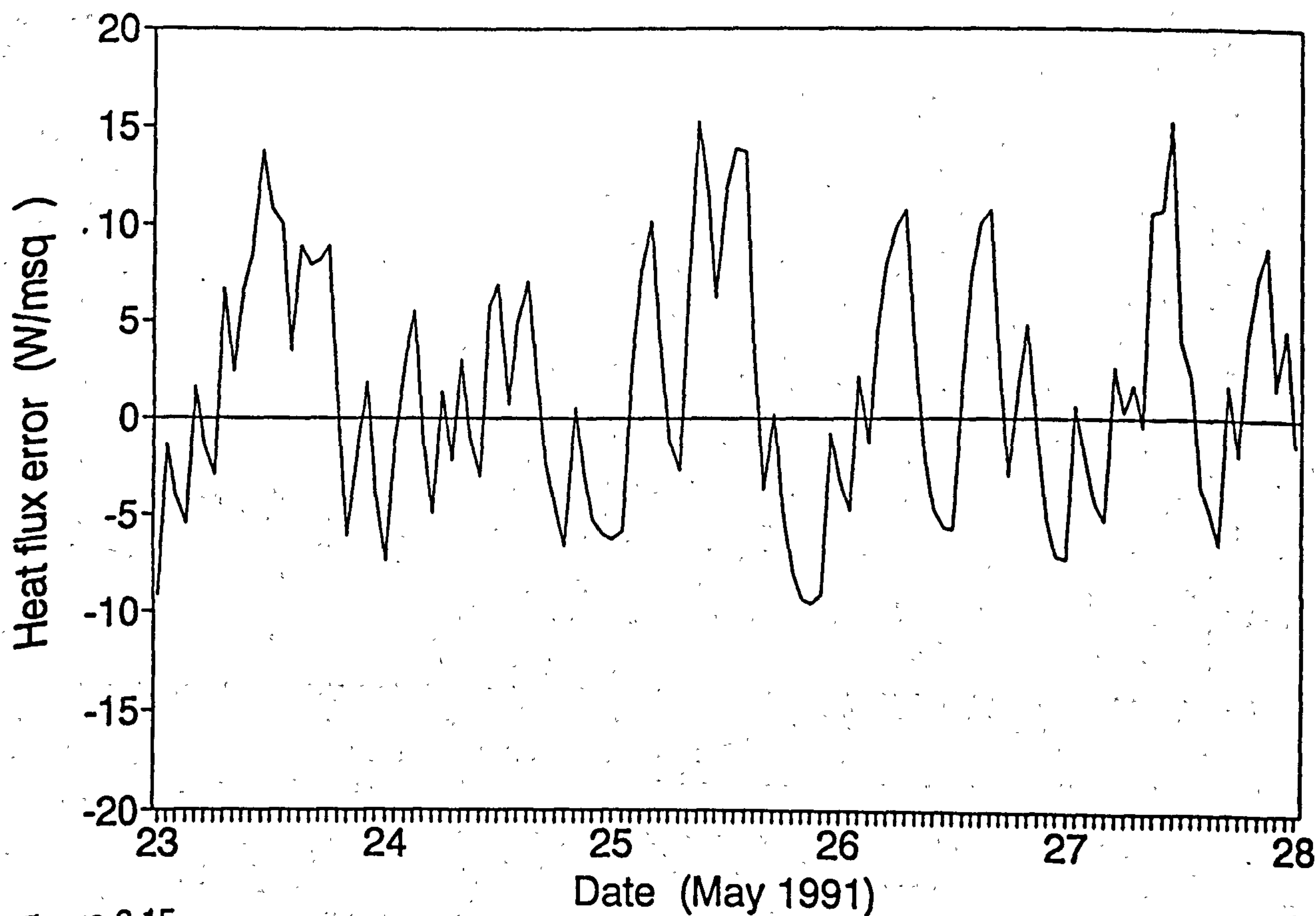


Figure 6.15

Relative contributions to zone temperature error

Double glazed test rooms

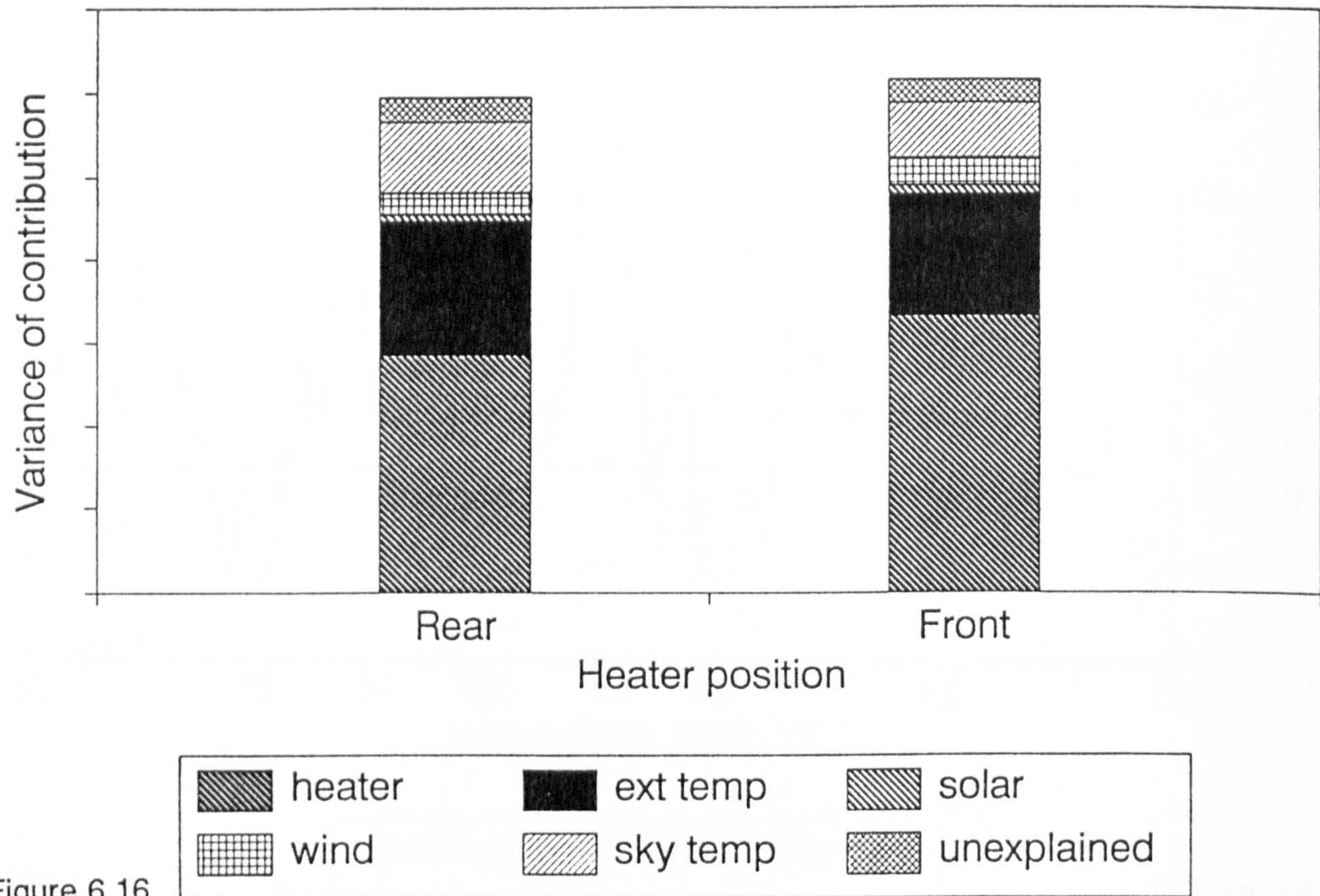


Figure 6.16

Zone temperature prediction error responses

Double glazed room with heat source at rear

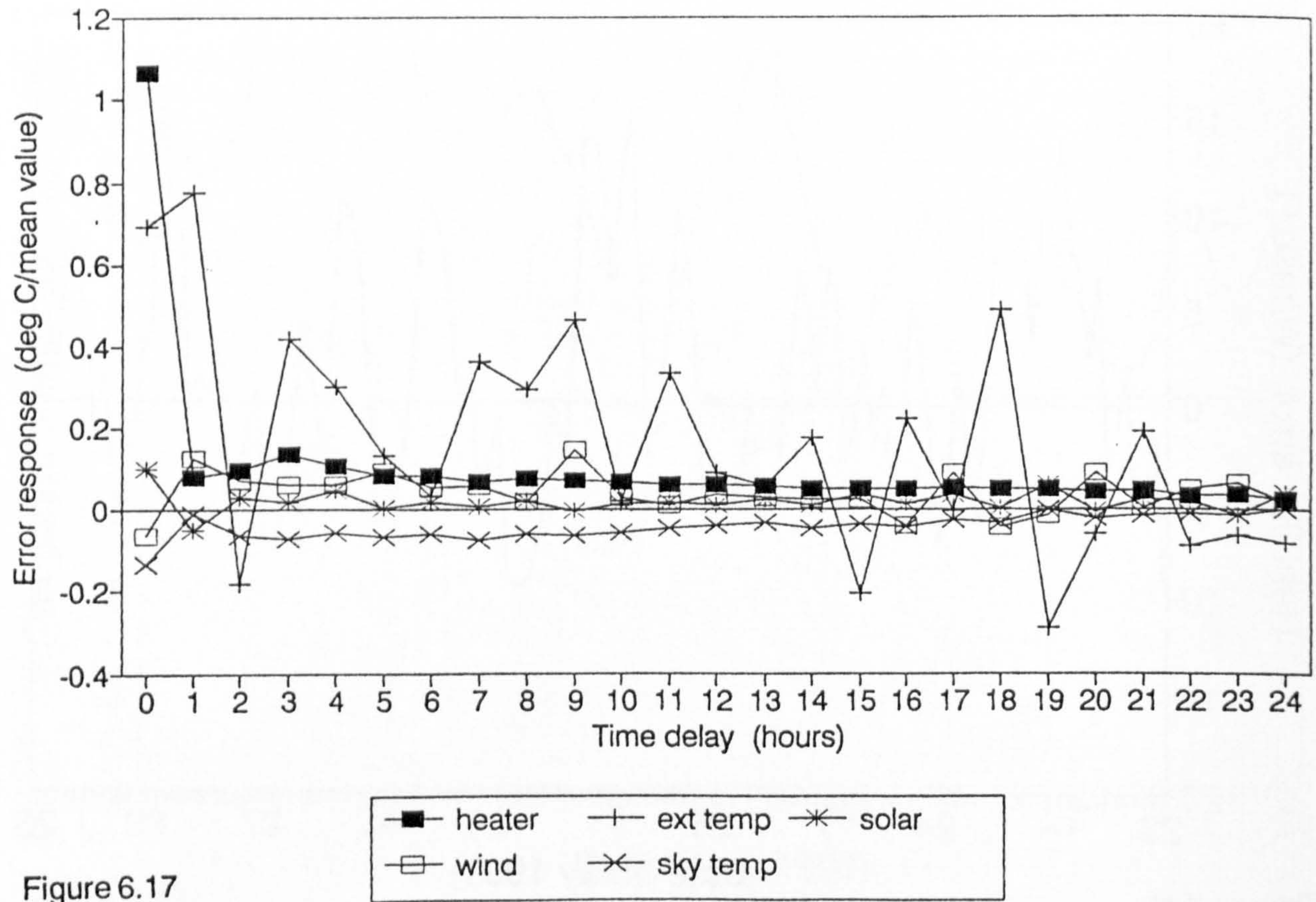


Figure 6.17

Zone temperature prediction error responses

Double glazed room with heat source at front

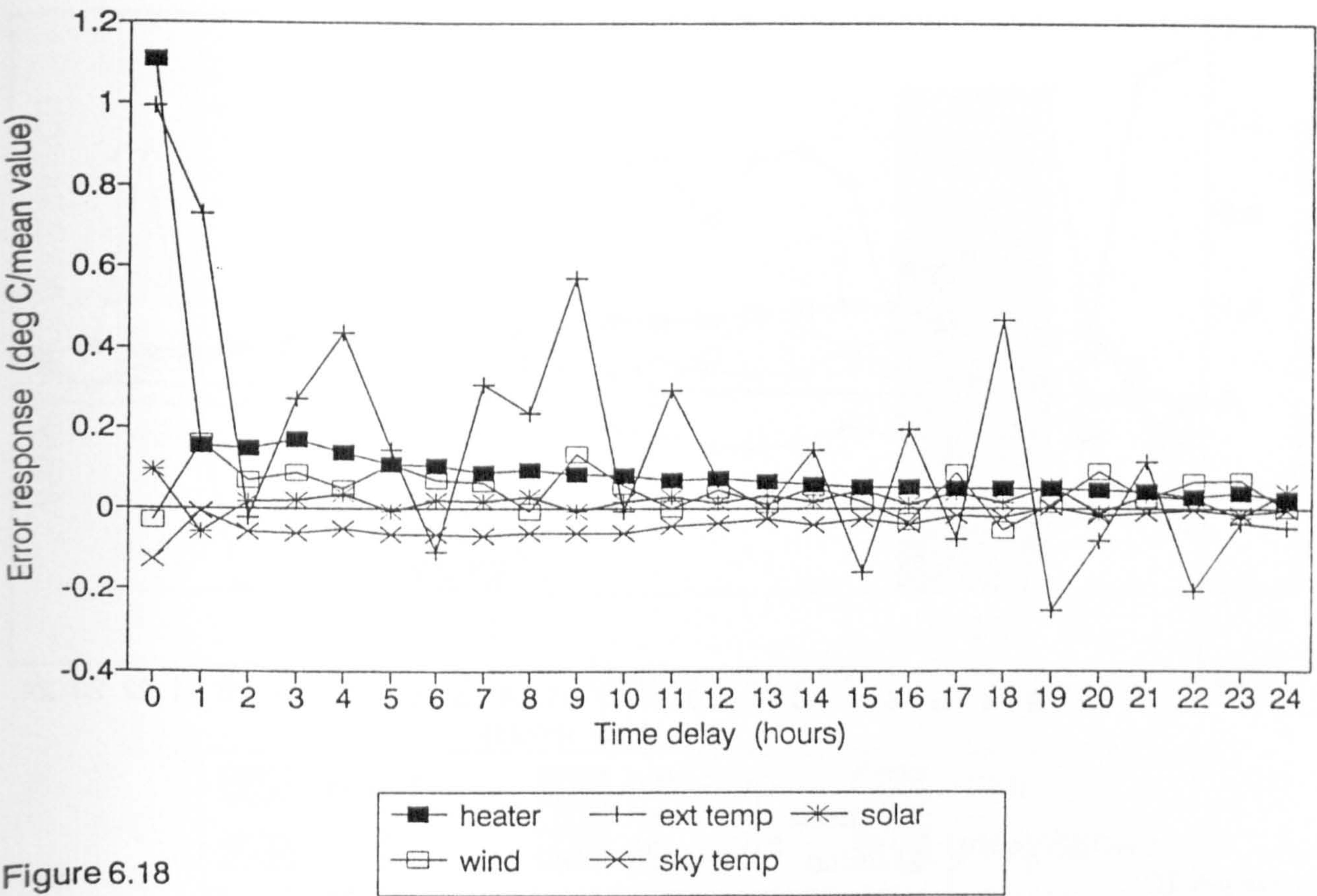


Figure 6.18

Relative contributions to floor temperature error

Double glazed test rooms

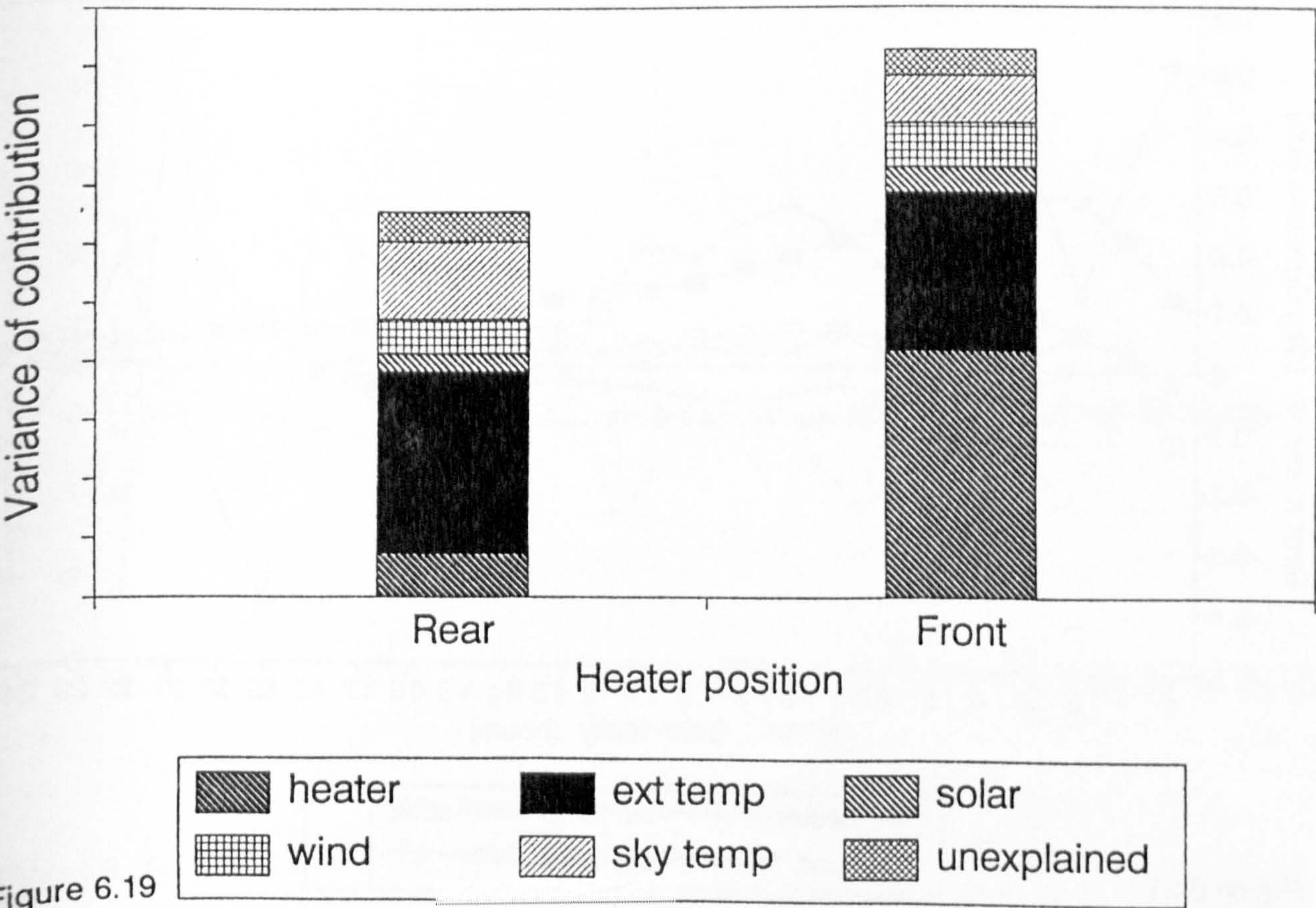


Figure 6.19

Floor temperature prediction error responses

Double glazed room with heat source at rear

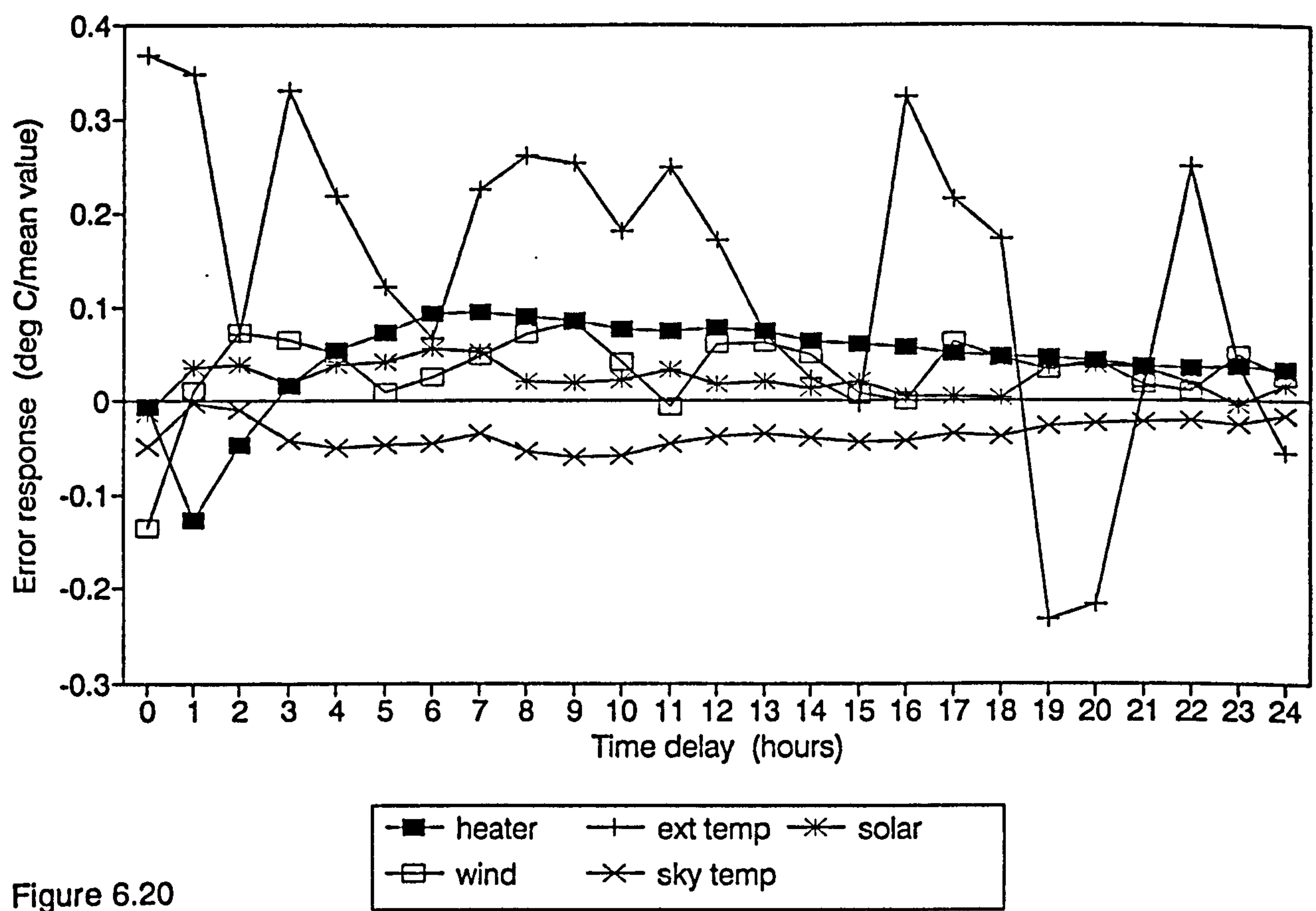


Figure 6.20

Floor temperature prediction error responses

Double glazed room with heat source at front

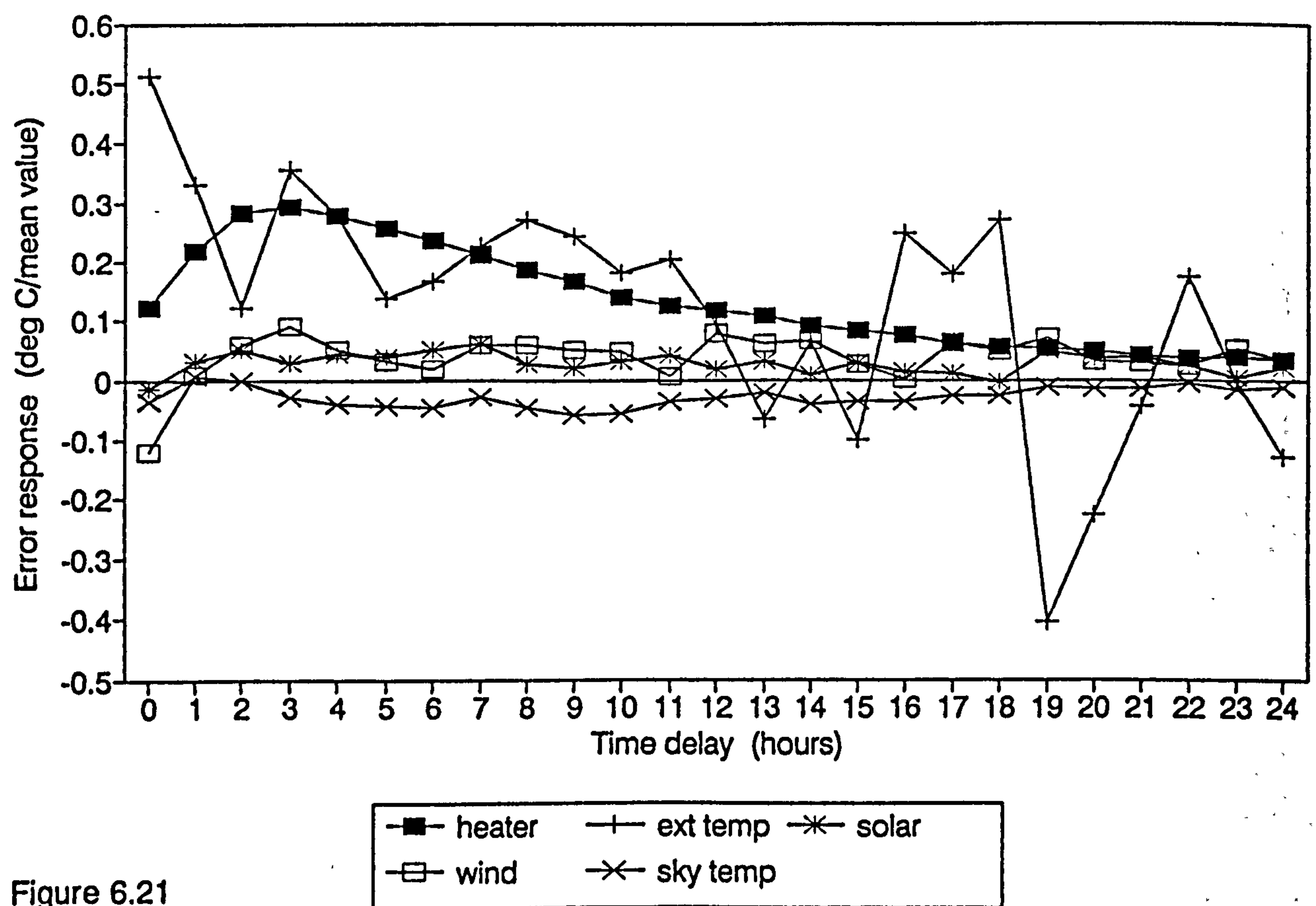


Figure 6.21

Relative contributions to floor heat flux error

Double glazed test rooms

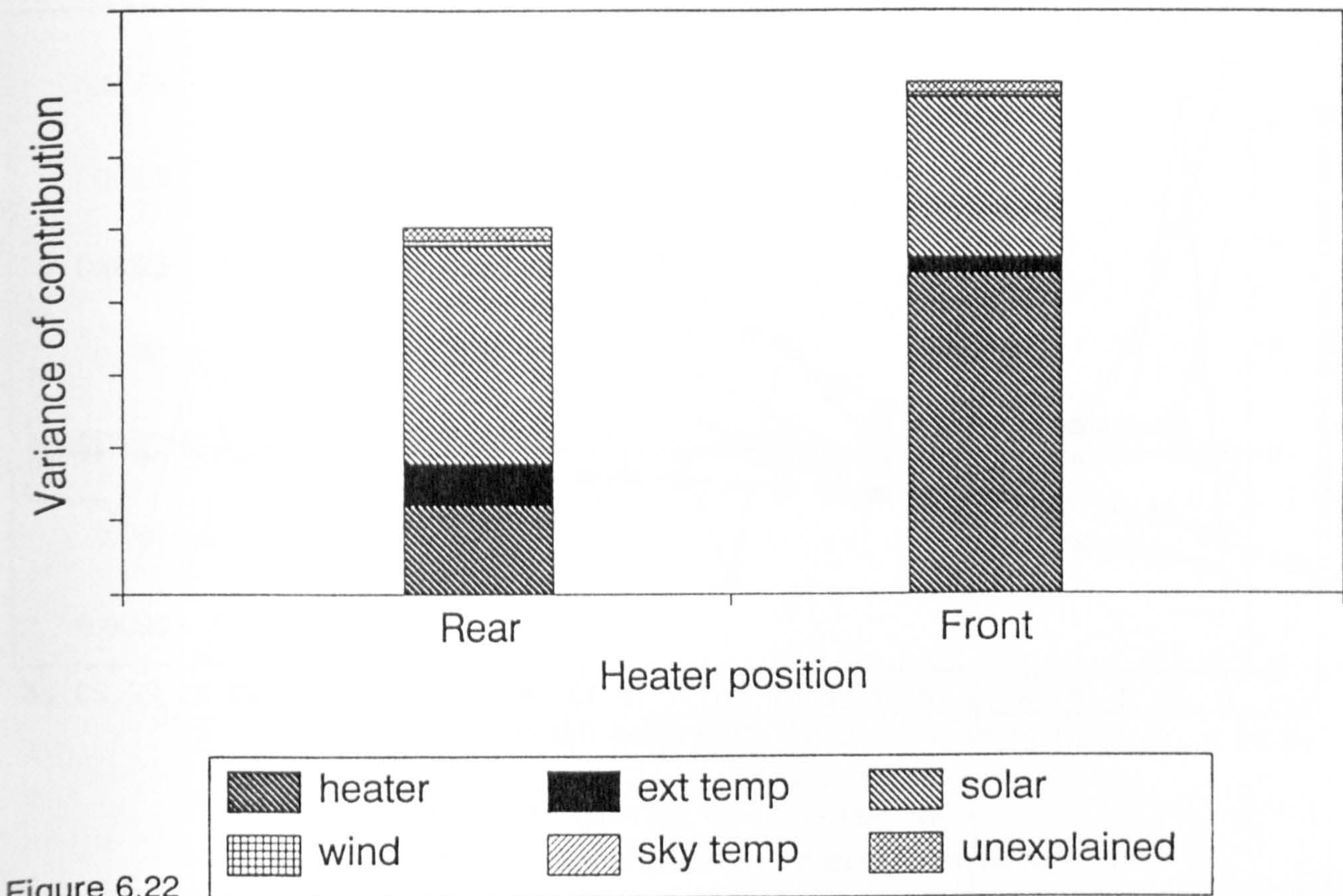


Figure 6.22

Floor heat flux prediction error responses

Double glazed room with heat source at rear

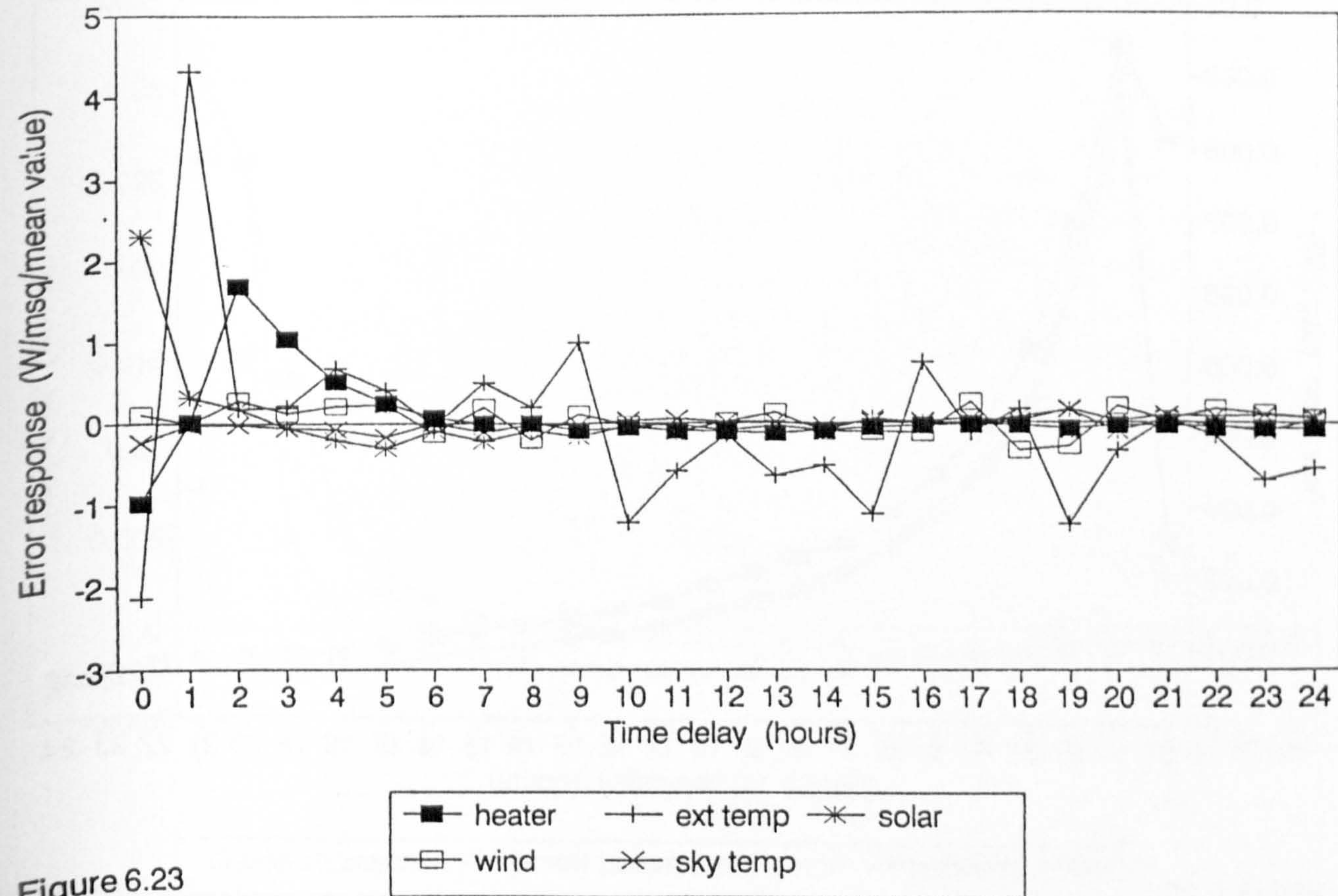
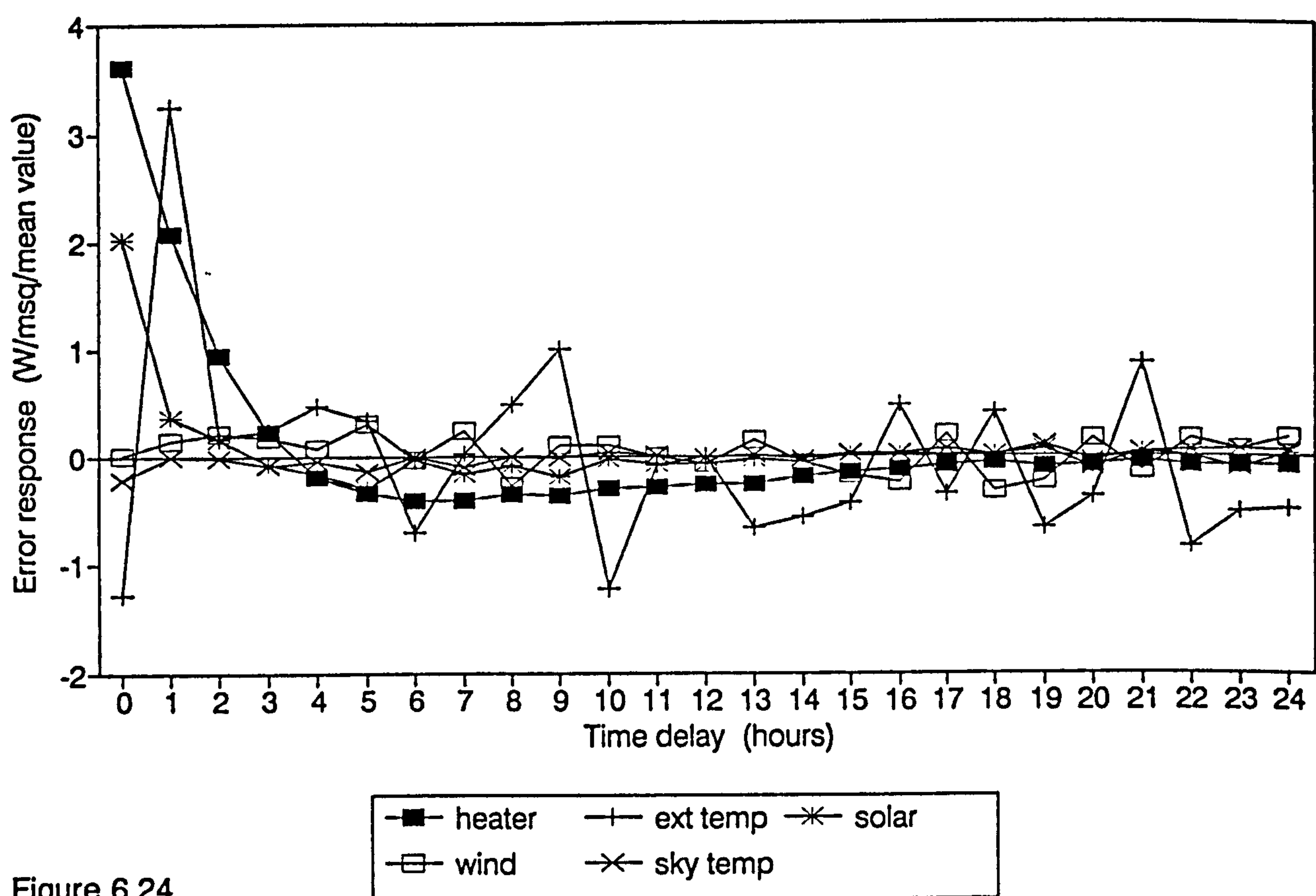


Figure 6.23

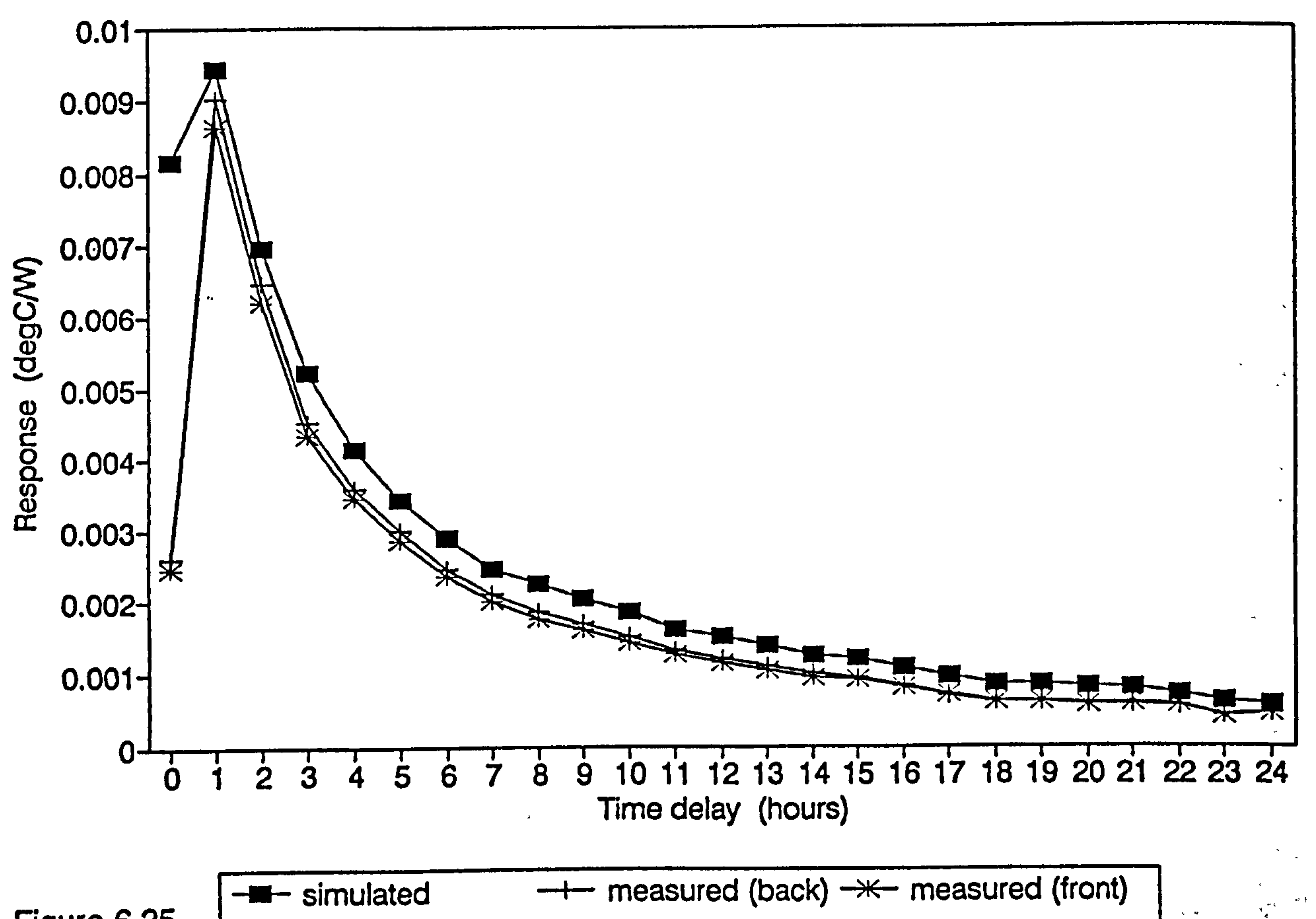
Floor heat flux prediction error responses

Double glazed room with heat source at front

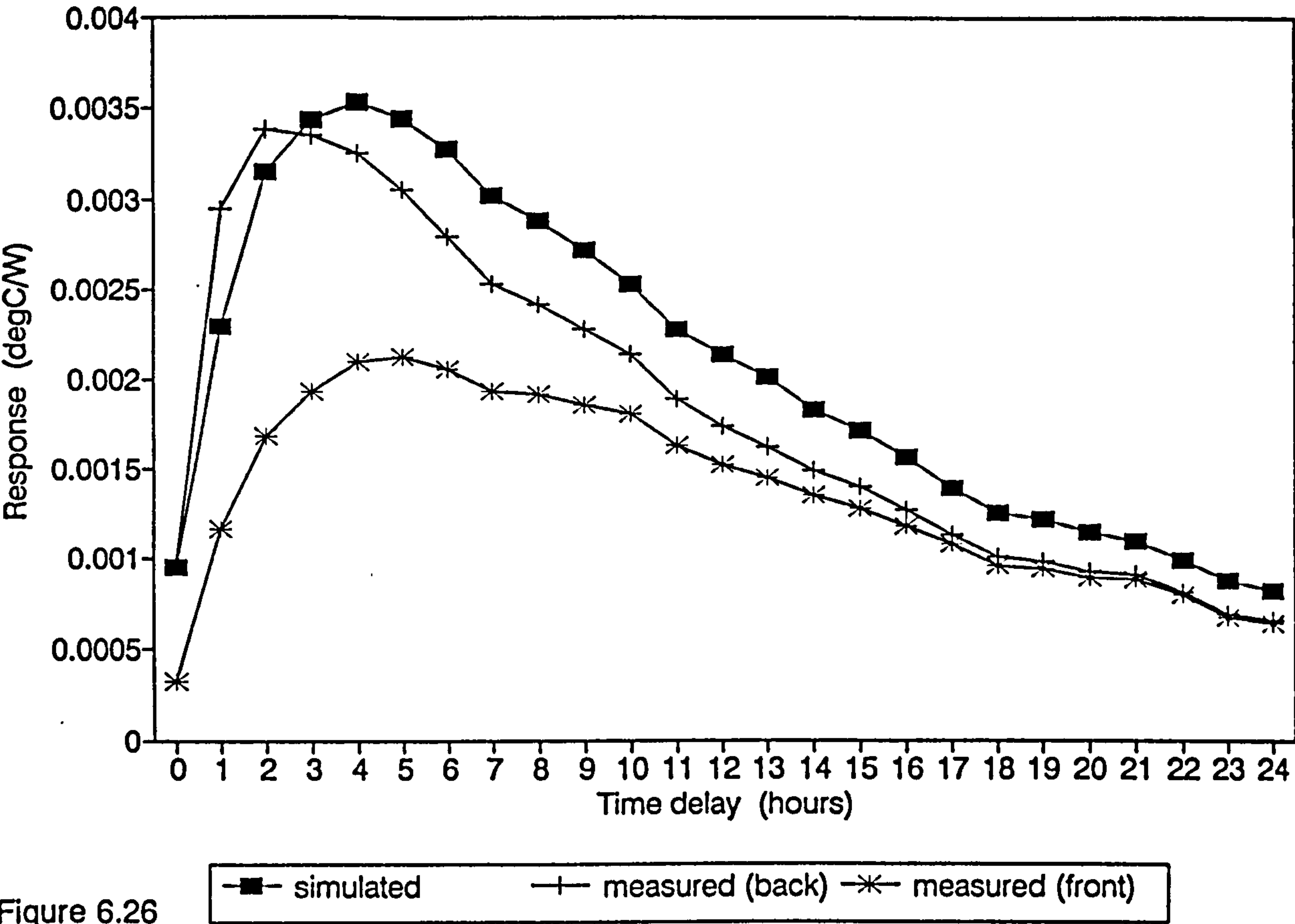


Simulated and measured responses to heater power

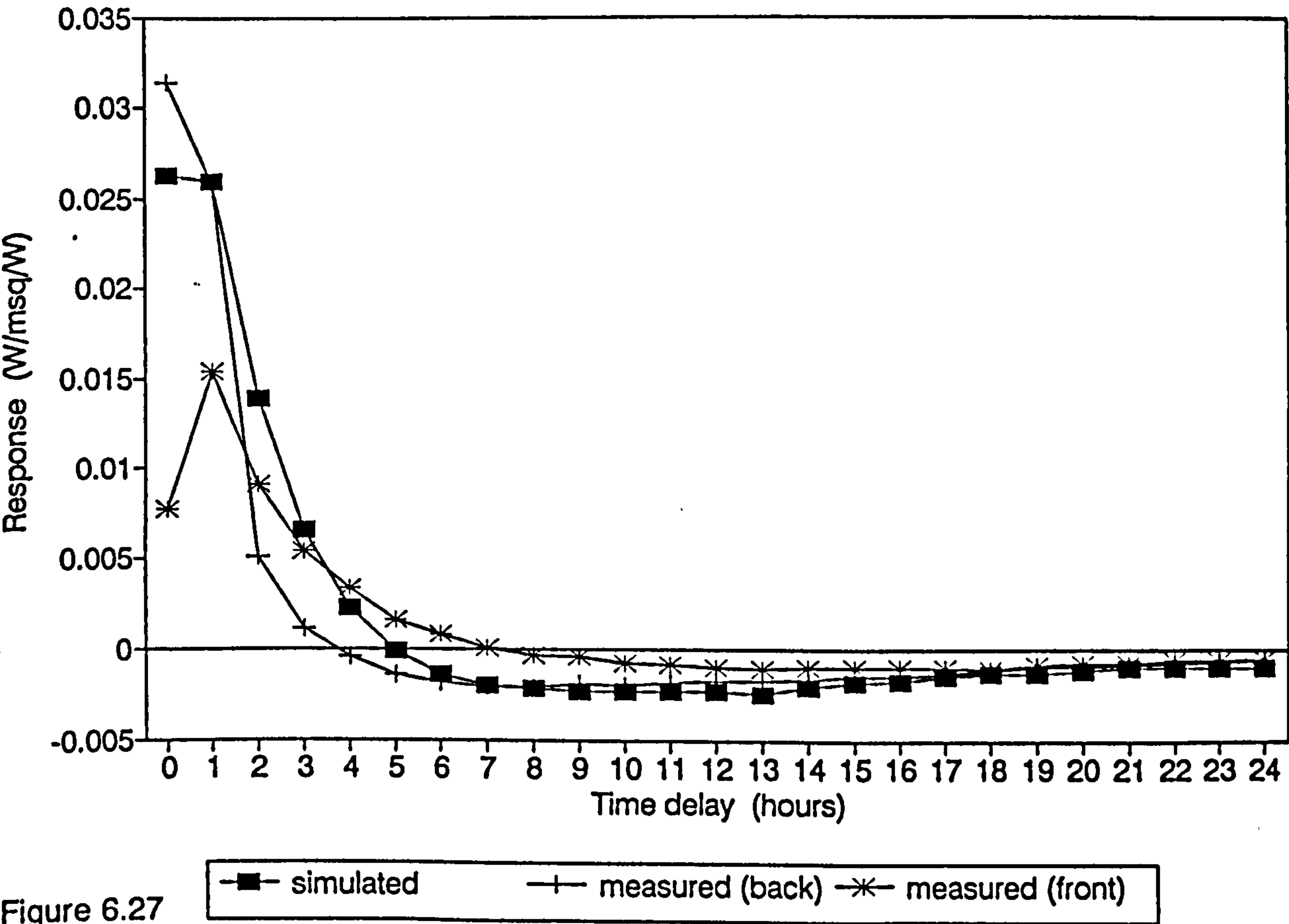
Zone temperature response



Simulated and measured responses to heater power
Floor temperature response



Simulated and measured responses to heater power
Floor heat flux response



CHAPTER 7

CONCLUSIONS

This thesis began by asserting the need for model testing. A review of model validation has lead to the development of a theoretical framework within which validation experiments can be conducted. This centres around the division of the model testing process into the twin activities of verification and validation, and the notions of context and of falsifiability.

The testing process has been divided into two distinct stages, verification and validation. Verification consists of determining that the implementation of a simulation model (most frequently a computer program) correctly carries out the calculations intended. Validation consists of the altogether more difficult task of determining whether the simulation model produces predictions which adequately represent physical reality. It has been concluded that empirical validation, in which the predictions of a model are compared with measured results, forms a central part of the model testing process. There are relatively few thermal processes in buildings which are so well understood that definitive models exist against which a new model can be tested. For all the remaining mechanisms, comparisons between measured data and the predictions of the new model represent the only way of establishing its likely accuracy in use.

It has further been concluded that model validation must always be context sensitive: it must take into account the use to which a model is to be put. Indeed, it has even been argued by some modellers that the whole process of model development, including validation, should be driven by a detailed knowledge of the ultimate application of the model.

It has been asserted that the process of model validation is, in principle, no different to the process of validating any other scientific theory. This leads to a view of validation as the process of successively failing to falsify a model. As this continues the area of applicability, over which the model can be used with confidence, progressively expands. When a validation test is encountered which the model fails, that area is delineated. Refinement of the model is then required to allow it to pass that particular test, and to expand further its area of applicability.

The tools available to the model validator have been reviewed. It has been shown that the overall process of model testing can be broken down by considering the three stages which make up the modelling activity: the formulation of a conceptual model, generation of a physical model and finally implementation of a computer model. Each of these stages can be verified and validated independently by applying the appropriate testing tools.

An historical review of validation activities in the building thermal modelling field has charted the emergence of a consistent approach to the empirical validation of building thermal simulation models, which takes account of the distinction between validation and verification, and of the notions of context and falsifiability. It has also revealed that not all subsequent validation exercises have followed this approach, and has indicated some of the problems which have been encountered as a result.

The analytical tools available to the empirical validator have been reviewed, and it has been concluded that there is a wealth of criteria for determining the validity of a model from the results of such comparisons. All of these make different assumptions about the nature of discrepancies between simulation and reality, and embody different criteria for validity. Not surprisingly they may give different results when applied to a given set of measured data and model predictions. A model which is deemed valid using one criterion may be deemed invalid by another. The task of selecting the criterion most appropriate in a given situation lies with the validator.

Whilst there has been found to be no shortage of methods to declare whether a model is valid, there is a distinct lack of tools which allow the reasons for discrepancies between model predictions and measured data to be identified. This is perhaps the most important function of empirical validation: to point to the areas towards which future model development should be directed. Existing tools have been found to be largely based on 'black box' statistical methods, and to rely on the use of cross correlation or spectral analysis techniques.

A new tool has been presented which differs from previous techniques in that it attempts to model the simulation error process explicitly, in terms of the variables which are driving both the simulation and the real world experiment. After a review and critical appraisal of potential modelling techniques an impulse response representation of the multi-variate linear time invariant system was selected. The method has been implemented, and tested in two stages. First the implementation of model fitting process was tested (or verified) by generating data sequences from a series of inputs and known impulse responses. The method proved capable of recovering the impulse responses from the input and output records. This part of the testing process provided a number of additional insights into how best to design experiments for the identification of impulse response models:

- the quality of response identification depends very strongly on the characteristics of the variables driving the test building. In particular, identification is much harder in the face of slowly moving, or strongly auto-correlated excitation. Obviously the nature of some driving forces (for example climate variables) is not generally under the control of the experimenter, but whenever possible driving forces should be chosen to have minimum auto-correlations. Randomising a variable is one way of achieving this. Steady driving forces (for example a heater operated at constant power) should be avoided at all costs.
- cross-correlations between the forces driving the test building reduce the quality of identification. Once again, in some cases these interactions will be unavoidable, but wherever possible inputs should be chosen not to correlate with each other, and not to correlate with other, predetermined, inputs. Again, randomisation is one way of achieving this.
- excessively large entries in the tail of a response may indicate that the identification process is not considering sufficiently long time delays. The identification should be repeated including more response terms.

The second stage in the testing process set out to determine whether the error modelling technique could provide useful insight into the reasons for discrepancies between simulation results and reality. To this end the building thermal simulation program SERI-RES was used

to generate a quasi-truth data set. A series of modified simulation results was then generated by making small perturbations to selected model input parameters. The magnitude of these perturbations was such that the resulting simulation 'errors' would not have been detected by graphical inspection of the results, or by an output uncertainty based test of validity. The new method, however, has proved capable of identifying the location of the altered parameter within the model, and examining the dynamic characteristics of the resulting error model has allowed the nature of the parameter to be deduced. Furthermore, a series of conclusions about the design of model validation trials and the subsequent identification and interpretation of error impulse responses have been drawn. When designing a validation experiment:

- surface heat flux data is generally a more sensitive indicator of the source of simulation errors than the corresponding surface temperature, and thus measuring heat fluxes will yield a more useful data set, and
- some processes (for example glazing heat loss and ventilation) may be very difficult to separate. Wherever possible, one of the processes should be 'designed out' of a validation experiment. For example, in an experiment designed to assess the quality of glazing heat loss modelling, ventilation should either be extremely accurately measured and modelled, or eliminated. If the latter course is chosen this will of course affect the context of the validation experiment: the model will not have been tested for situations where ventilation is present.

When interpreting the resulting error impulse responses:

- oscillation at the beginning of a response may indicate that the variable driving that response does not contain sufficient high frequency variation to allow the response at short time delays to be identified, and
- it is important that the interpretation process first considers the relative contributions to reconstituted error as well as at the impulse responses themselves. An apparently large impulse response does not necessarily imply a significant contribution to the error. Once a significant contribution has been established, however, examining the appropriate impulse response can yield useful information about the error process dynamics.

It has been argued that outdoor test rooms represent a powerful vehicle in which to carry out empirical model validation. They are of realistic size and construction, and are exposed to real climate. At the same time they offer sufficient control of experimental uncertainty to allow meaningful interpretation of comparisons between measured data and simulation results. In view of these arguments the new tool has been applied to the problem of comparing the building thermal simulation model SERI-RES with measurements from a pair of outdoor test rooms. An empirical validation experiment has been designed and carried out in a pair of outdoor test rooms. The experiment featured a pair of such test rooms. These were double glazed, and heated by panel radiators. In one room the radiator was placed beneath the window, in the other it was on the opposite wall. The analytical test of the new analysis technique previously presented had revealed that the tool was most powerful when identifying the response to randomised inputs. In the case of an outdoor test facility the climatic variables are beyond the control of the experimenter, but other inputs which determine the performance of the room can be chosen to maximise the value of the experiment. The most obvious of these is heater operation. For the experiment described the heaters were therefore operated to a pseudo-random schedule.

The performance of the test rooms was predicted using the simulation program SERI-RES. The simulations were carried out 'blind' by using a description of the test building which had been prepared prior to the data collection, ensuring that there was no possibility of the representation of the building within the model being influenced by a knowledge of the required results. The simulations were conducted by supplying to the model, on an hour by hour basis, the power input to the rooms and climate data which had been measured during the experiment. The model was used to predict the test room zone temperature, surface temperatures and surface heat fluxes for comparison with the corresponding measurements.

Preliminary comparisons between simulation results and measured data revealed that the model tended to overpredict temperatures and heat fluxes in both of the test rooms. The degree of overprediction was less severe in the room with the heater at the back than when the heater was

located beneath the window. It has been suggested that this may be due to the fact that in the former case a convective loop forms between the heater and window which promotes better mixing of the air within the room, bringing its performance more closely into line with the assumption of uniform air temperature in the model.

The new technique was then applied to the simulation errors for both rooms, and the relative contributions to the error from the various forces driving the performance of the rooms determined. In both cases it was found that the linear time-invariant error model could explain the bulk of the observed error. It was seen that for the room with the heater at the front (beneath the window) the predominant source of air temperature prediction error was the operation of the room heater. When the heater was placed at the rear, however, the contribution from this source was much smaller, resulting in the smaller overall prediction errors noted previously.

For both rooms the air temperature prediction error response was instantaneous: an error appeared as soon as the heater was operated. When other quantities were examined the new technique proved capable of revealing subtle dynamic differences in the error structures for the two rooms. The error in predicted floor temperature is consistently positive (indicating overprediction) when the heat source is located at the front of the room. With the heat source at the back it is initially zero, becomes negative one hour after the heater is operated and then becomes positive. At this point it decays in the same way as for the room with the heater at the front. When floor heat flux is examined a similar story emerges. With the heat source at the front of the room the initial overprediction decays over the first few hours to produce subsequent underprediction. With the heater at the rear of the room, however, the initial error is negative, followed by overprediction which decays over the following few hours. As with floor temperature, the overall error is smaller than when the heater is placed at the front of the room.

It is clear from the above discussion that the highly simplified treatment of the room air and

heat source within SERI-RES can be the source of significant prediction errors, in particular when the heater is positioned below the room window, a common choice of location. Other simulation models which are currently in use employ a similar representation. A better model of how a heater interacts with a room could therefore produce a significant improvement in the quality of the predictions of these models. In particular, such an improved model should take into account the position of the heater in relation to the room glazing.

The new tool has been used to demonstrate that by far the most significant source of prediction errors in SERI-RES is the operation of the test room auxiliary heat source. The technique has proved sufficiently sensitive to detect the change in error structure which occurs as the position of the heat source within the test room is changed. At present very few thermal models treat the processes within buildings at a level of detail which allows the coupling of the heat source to the room air to be treated at the level of detail required to eliminate these problems. Such models are commonly used to predict the energy consumption of heated buildings, and in this context the errors identified will have a significant effect. It is therefore clear from the validation exercise presented here that future model development efforts should be directed towards this area.

The successful application of the new tool has thus been demonstrated on a comparison between the predictions of a thermal simulation model and measurements made in a simplified test building. This approach has the advantage that many of the phenomena present in real buildings can be eliminated from such test facilities, making the interpretation of results more straightforward. In justifying this choice of test building, however, a complete spectrum of facilities was identified, ranging from laboratory rigs, through simplified and unoccupied structures, to occupied buildings in normal operation. The new technique could find application in the analysis of comparisons with data from any of these facilities. Its particular value, however, will lie in the interpretation of data from outdoor facilities, where many inter-correlated meteorological variables may have a complex effect on the experiment, and individual driving forces cannot generally be eliminated as part of the experimental design.

The technique as described has used a linear time-invariant model to represent the discrepancies observed between simulations and reality. For almost all of the cases examined, both numerically and with empirical data, this model has proved capable of explaining the vast majority of the simulation error in terms of the model driving forces. Should cases arise where this is not possible, the impulse response model can be extended to include higher order, non-linear, terms or time varying effects.

In the course of the testing process a further area in which the new tool could be enhanced has been identified. The interpretation of the outputs of the tool, which take the form of multiple impulse responses, currently requires a certain amount of skill and experience. Further work might seek to assist in this interpretation by incorporating the knowledge gained from numerical trials such as those described in Chapter 5 into an expert system. Such a system could then form a powerful aid to the interpretation of the outputs of the existing tool.

The work which has been presented in this thesis has concentrated on the validation of models which simulate the thermal performance of buildings. However, the tool which has been developed has the potential for application to the validation of any dynamic simulation model. In the field of building simulation it might find application in the comparison of measurements with numerical models of airflow or the performance of heating or cooling plant. Moving slightly further from the original application, the method might be used in the validation of models of the propagation of stresses through structures. In the mechanical engineering field such models are of great interest. They have obvious applications in structural design, and are used throughout the automotive industry as well as in construction. There is now additional interest in using them as instruments with which to analyse the results of non-destructive testing.

Returning for a moment to the automotive industry, there is now considerable work underway in the field of dynamic modelling of engine performance, with the aim of obtaining more efficient engine control, with reduced emissions [Amstutz and Del Re, 1995]. Such models, of

course, require validation before use, and this validation is currently carried out using the type of graphical comparison techniques which were common in the building simulation field in the 1970s. The new technique could obviously find application in the more detailed validation of such models.

The technique developed here could be applied to larger models than the building thermal simulations considered here, such as climatological prediction systems. In this case both the intrinsic non-linearity of the systems being represented and the sheer size of the computational problem would probably become limitations. The former could be resolved by incorporating higher order, non-linear, terms in the impulse response error model, as briefly described in Chapter 3. The latter could be addressed by exploring some of the computationally more efficient solutions, discussed briefly in Chapter 3 but rejected as unnecessary for the present application.

The handful of possible applications for the new tool which have been listed above indicate the vast range of applications over which it could, in principle, find application. In fact, of course, it has the potential for use in any areas where a dynamic model is to be validated by comparison with measured data. In such alternative applications the physical processes underlying the modelling problem might not be as well understood as they are for buildings, and repetition of the testing process described in Chapter 5 would provide a valuable way of developing some insight into how best the resulting error models should be interpreted in new situations.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ackoff R L. 1962. *Scientific Method: Optimising Research Decisions*. Wiley, New York.

Alamdari F and Hammond G. 1983. Improved Data Correlations for Buoyancy Driven Convection in Rooms. *Building Services Engineering Research and Technology*. 4(3). p 106-112.

Allen E J. 1984. A comparison of the predictions made by two complex thermal simulation computer programs, for simple test cases, performed under IEA Task 8 (Passive and Hybrid Solar Low Energy Buildings). BRE Note: 168/84.

Allen E J. 1985. Thermal model validation. BRE Outline paper.

Allen J P and Whittle J K. 1986. The appropriateness of heat transfer coefficient algorithms. *Proceedings CIBSE Fifth International Symposium on the use of computers for environmental engineering relating to buildings*. Bath, UK.

Amstutz A and Del Re L R. 1995. EGO sensor based robust output control of EGR in deisel engines. *IEEE Transactions on Control Systems Technology*. 3(1). p39-48.

Anand D K, Kennish W J, Knasel T M and Stolarz A C. 1979. Validation methodology for solar heating and cooling systems. *Energy*. 4(4). p 549-560.

Anscomb P J. 1967. Topics in the investigation of linear relations fitted by the method of least squares. *Journal Royal Statistical Society*. B29. p 1-29.

Arumi-Noe F and Northrup D O. 1979. A field validation of the thermal performance of a passively heated building as simulated by the DEROB system. *Energy and Buildings*. 2(1). p 65-75.

Atkinson B A, Barnaby C S, Wexler A H and Wilcox B A. 1981. Validation of CALPAS3 computer simulation program. *Progress in Passive Solar Energy*. 6. p 358-361.

Balci O and Sargent R G. 1980. Bibliography on Validation of Simulation Models. *Newsletter of the TIMS College on Simulation and Gaming*. 4(2).

Baleynaud J M, Petit M and Trombe A. 1991. Attached Sunspace - Sensitivity Factors. *Solar Energy*. 46(3). p 149-166.

Basler and Hofmann. 1985. Simulation model validation using measured data from the Swiss Trombe wall test cell. Zurich, Switzerland.

Bauman F, Andersson B, Carroll W L, Kammerud R and Friedman N E. 1983. Verification of BLAST by comparison with measurements of a solar dominated test cell and a thermally massive building. *Journal of Solar Energy Engineering*. 105. p 207-216.

Bendat J S and Piersol A G. 1986. Random Data Analysis and Measurement Procedures. John Wiley and Sons Inc.

Bierman G J. 1977. Factorisation Methods for Discrete Sequential Estimation. Academic Press.

Bland B H and Bloomfield D P. 1986. Validation of conduction algorithms in dynamic thermal models. *Proceedings CIBSE Fifth International Symposium on the use of computers for environmental engineering relating to buildings*. Bath, UK.

Bland B H. 1992. Proposals for further work on testing in IEA21. International Energy Agency Research Note: IEA21RN298/92.

Bland B H. 1992a. Development of tests to evaluate the treatment of solar radiation in the prediction of building performance. BRE Note: N56/92.

Bland B H. Conduction in dynamic thermal models: Analytical tests for validation. *Building Services Engineering Research and Technology*. 13(4). p 197-208.

Bloomfield D P. 1985. Appraisal Techniques for Methods of Calculating the Thermal Performance of Buildings. *Building Services Engineering Research and Technology*. 6(1). p 13-20.

Bloomfield D P, Brewer R, Stringer B, Cooper I and Lera S. 1988. IEA Task VIII - The UK experience. Improving the design of low energy passive solar houses. *Building Research and Practice*. No. 6.

Bloomfield D P, Lomas K J and Martin C J. 1992. Validation of Programs for Calculating the Thermal Performance of Buildings. BRE Information Paper IP7/92.

Bloomfield D P, Eppel H, Lomas K J and Martin C J. 1993. Multi-national empirical validation using high quality datasets. *CIBSE 1993 Annual Conference*. Manchester.

Bohlin T. 1987. Model Validation. *Encyclopedia of Systems and Control* edited by: Singh M G. Pergamon Press.

Bowman N T. 1985. Building Energy Evaluation Techniques Assessed. *Building*. 21(6). p 80-81.

Bowman N T and Lomas K J. 1985. Empirical Validation of Dynamic Thermal Computer Models of Buildings. *Building Services Engineering Research and Technology*. 6(4). p 153-162.

Bowman N T and Lomas K J. 1987. Modifications to Basic Cell Description for SERI-RES Simulations. Leicester Polytechnic report to Energy Monitoring Company Ltd.

Box G E P and Jenkins G M. 1976. Time Series Analysis: Forecasting and Control. Holden Day.

Box G E P, Hunter W G and Hunter J S. 1978. Statistics for experimenters: An introduction to Design, Data Analysis and Model Building. Wiley, New York. 202 c 97 667 SW5

Brigham E O. 1974. The Fast Fourier Transform. Prentice Hall.

Burch D M, Peavy B A and Powell F J. 1975. Comparison between measured-computer predicted hourly heating and cooling energy requirements for an instrumented wood framed townhouse subjected to laboratory tests. *ASHRAE Transactions*. 2.

Butterfield M H. 1981. A method of model verification for large perturbation transients with examples from nuclear power. *Proceedings of IEE conference*. Warwick, UK.

Butterfield M H and Thomas P J. 1983. Methods of Quantitative Model Validation for Linear and Non-linear Systems, with Applications to Nuclear Power Plant. Winfrith. UKAEA, Dorchester, DORSET.

Butterfield M H and Thomas P J. 1986. Methods of quantitative validation for dynamic simulation models. Part 1: Theory. *Transactions of the Institute of Measurement and Control*. 8(4). p 182-200.

Butterfield M H and Thomas P J. 1986. Methods of quantitative validation for dynamic simulation models. Part 2: Examples. *Transactions of the Institute of Measurement and Control*. 8(4). p 201-219.

Candau Y and Piar G. 1991. An application of spectral decomposition to model validation in building thermal analysis.

Canha de Piedade A, Moret Rodrigues A and Pedro des Fritas J. 1986. The use and validation of simulation computer programs for passive solar building design in Portugal. *Proceedings CIBSE Fifth International Symposium on the use of computers for environmental engineering relating to buildings*. Bath, UK.

Carter C. 1979. A validation of the model expansion method of modelling heat conduction in passive solar buildings. *Solar Energy*. 23(6). p 537-539.

Caswell H. 1976. The validation problem. *Systems analysis and simulation in Ecology* edited by: Patten B C. 4. p 313-325. Academic Press, New York.

Chantret A. 1982. Experimental validation of heating consumption models. *Promoclim E*. 13(E5). p 259-268.

Chantret A. 1983. Experimental validation of simplified models of heating energy consumption. *Annales de l'Institut Technique du Batiment et des Travaux Publics*. 416. p 69-75.

- Chartered Institution of Building Services Engineers (CIBSE). 1986. CIBSE Guide A3: Thermal Properties of Building Structures.
- Chatfield C. 1975. Analysis of Time Series: Theory and Practice. Chapman and Hall.
- Churchman C W, Ackoff R L and Arnoff E L. 1957. Introduction to Operations Research. Wiley, New York.
- Clarke J A and Forrest I. 1978. Validation of the ESP thermal simulation program. ABACUS occasional paper 61. University of Strathclyde.
- Clarke J A. 1985. Energy simulation in building design. Adam Hilger Ltd.
- Cockcroft J P. 1982. Validation of buildings and systems energy prediction using real measurements. *Computer Aided Design*. 14(1). p 39-43.
- Cohen K J and Cyert R M. 1961. Computer Models in Dynamic Economics. *The Quarterly Journal of Economics*. LXXV. p 112-127.
- Commenges D. 1984. The Deconvolution Problem: Fast Algorithms Including the Preconditioned Conjugate-Gradient to Compute a MAP Estimator. *IEEE Transactions on Automatic Control*. AC-29. p 229-243.
- Cools C and Gicquel R. 1988. The project PASSYS: Solar Energy applications in buildings. CEC DGXII.
- Croome D J and Moseley P J. 1986. A comparison of temperature prediction methods for lightweight structures. *Proceedings CIBSE Fifth International Symposium on the use of computers for environmental engineering relating to buildings*. Bath, UK.
- Cuénod M and Durling A. 1969. A discrete time approach for systems analysis. Academic Press, New York.
- Davis J C. 1986. Statistics and data analysis in Geology. Wiley. New York.
- Davis M G. 1975. A thermal model for a solar heated building. *Building Science*. p 67-76.

Day B. 1982. Computation of the dynamic thermal performance of buildings: Science Research Council's specially promoted program. *Computer Aided Design*. 14(1). p 49-54.

Dewson T, Day B and Irving A D. 1992. Investigation of techniques to extract building system performance assessment characteristics from field data. Final report on SERC Grant no. GR/F/07057.

Dorny C N. 1980. A Vector Space Approach to Models and Optimisation. John Wiley and Sons.

Emery A F, Kippenham C J and Heerwagen D R. 1981. The simulation of building heat transfer for passive solar systems. *Energy and Buildings*. 3(4). p 287-294.

Erhorn H, Gertis K, Rath J and Wagner J. 1987. Do computer calculations of the thermal performance of buildings agree with practical measurements ? Fraunhofer Institute report WB 9/87.

Fishman G S and Kiviat P J. The Analysis of Simulation Generated Time Series. *Management Science*. 13(7). p 525-557.

von Foerster H, Mora P and Amiot L W. 1960. Doomsday: Friday 13 November, AD 2026. *Science*. 132. p1291-1295.

Forsythe G E and Moler C B. 1967. Computer solution of Linear Algebraic systems. Prentice Hall.

Gelb A. 1974. Applied Optimal Estimation. The MIT Press.

Gentil S and Blake G. 1981. Validation of complex ecosystems models. *Ecological Modelling*. 14. p 21-38.

Golub G. 1965. Numerical methods for solving least squares problems. *Numer. Math*. 7. p 206-216.

Gough M. 1984. Analytical tests for use in validating dynamic thermal models - Part 1. BRE Note: N172/84.

Gough M. 1984. Analytical tests for use in validating dynamic thermal models - Part 2. BRE Note: N181/84.

Gough M. 1984a. IEA Task 8 - Validation report: Modelling techniques. BRE Note: N183/84.

Grelat A, Fauconnier R and Marachal J C. 1987. Numerical simulation of the thermal behaviour of buildings. *Promoclim E*. 18(3). p 131-146.

Grimmer D P, McFarland R D and Balcomb J D. 1979. Initial experimental tests on the use of small passive solar test boxes to model the thermal performance of passively solar heated building designs. *Solar Energy*. 22(4). p 351-354.

Haves P H. 1985. Modelling Different Forms of Heat Input using SERI-RES. Polytechnic of Central London Research in Building Group.

Haves, P H and Littler J. 1987. Refinements to SERI-RES. Polytechnic of Central London report to ETSU. S-1130.

Hearne J W. 1985. Sensitivity analysis of parameter combinations. *Applied Mathematical Modelling*. 9(2). p 106-108.

Herrero M A. 1986. An analysis of the SERI-RES simulation results of the UK base building used in the International Energy Agency Solar Task 8. BRE Note: N142/86.

Hitchin E R, Delaforce S R and Martin, C J. 1993. A comparison of the measured and simulated thermal response of a simple enclosure. *Building and Environment*. 28(2). p 189-199.

Holmes M J and Whittle G E. 1987. How accurate are the predictions of complex air movement models? *Building Services Engineering Research and Technology*. 8(2). p 29-31.

Hoover S V and Perry R F. 1989. Simulation: a problem solving approach. Addison Wesley.

van Horn R L. 1969. Validation. *The Design of Computer Simulation Experiments*. edited by: Naylor T H. p235-251.

van Horn R L. 1971. Validation of Simulation Results. *Management Science*. 17. p 247-258.

Howarth A T. The Prediction of Air Temperature Variations in Naturally Ventilated Rooms with Convective Heating. *Building Services Engineering Research and Technology*. 6(4). p 169-175.

Howell J R. 1973. Monte Carlo treatment of data uncertainties in thermal analysis. *Journal of Spacecraft*. 10(6). p 411-414.

Hsu D A and Hunter J S. 1977. Analysis of simulation-generated responses using autoregressive models. *Management Science*. 24(2). p 181-190.

Irving A D. 1988. Application of Statistical techniques to the validation of Multivariate Time Series Simulators. Rutherford Appleton Laboratory.

Irving A D. 1992. Stochastic Sensitivity Analysis. *Applied Mathematical Modelling*. 16. p 3-15.

Irving S J. 1982. Energy Program Validation. *Computer Aided Design*. 14(1). p 33-38.

Isermann R. 1981. Digital Control Systems. Springer-Verlag.

Ishida K and Udagawa M. 1991. Validation of the simulation method of room temperature and heat load with the long term measured data. *AIJ Journal of Architecture Planning and Environmental Engineering*. 430(12). p 23-31.

Jackson L B. 1986. Digital filters and signal processing. Kluwer Academic Publishers.

Jeandel A, Palero I and Laret L. 1991. An approach to thermal modelling and simulation of buildings at Gaz de France. *Proceedings Building Simulation '91*. p 539-546. Nice, France.

Jensen S O and van de Perre R. Tools for whole model validation of building simulation programs -experience from the CEC concerted action PASSYS. *Proceedings Building Simulation '91*. p 547-552. Nice, France.

Jong M T. 1982. Methods of discrete signal and system analysis. McGraw Hill.

Jørgensen S E. 1981. Model for lake Glumsø, Denmark. *Models for Water Quality Management*. edited by: Biswas A K. p 314-341. McGraw Hill Inc.

Judkoff R, Wortman D, O'Doherty B and Burch, J. 1983. A methodology for validating building energy analysis simulations. Solar Energy Research Institute report TR-254-1508.

Judkoff R D. 1986. International Energy Agency building simulation comparison and validation study. *Proceedings International Climatic Architecture Congress*. p 491-509.

Judkoff R. 1988. Validation of building energy analysis simulation programs at the Solar Energy Research Institute. *Energy and Buildings*. 10. p 221-229.

Kalema T. 1991. Thermal analysis of buildings - verification and further development of the TASE program. Tampere University of Technology Thermal Engineering Report.

Kalman R E. 1960. A new approach to linear filtering and prediction problems. *Journal Basic Engineering*. p 35-46.

Kruger P and Hallanger E C. 1974. Computer simulation and actual results of an energy conservation program. *ASHRAE Journal*. 16(10). p 52-60.

Kusuda T, Pierce E T and Bean J W. 1981. Comparison of calculated hourly cooling load and attic temperature with measured data for a Houston test house. *ASHRAE Transactions*.

Law A M and Kelton W D. 1982. Simulation Modelling and Analysis. McGraw-Hill.

Leggett R W and Williams L R. 1981. A reliability index for models. *Ecological Modelling*. 13. p 303-312.

Letherman K M, Palin C J and Park P M. 1982. The thermal frequency response of rooms. *Building and Environment*. 17(1). p 5-10.

Letherman K M, Palin C J and Park P M. 1982. The measurement of dynamic thermal response in rooms using pseudo-random binary sequences. *Building and Environment*. 17(1). p 11-16.

Levermore G. 1989. Which Program ? *Building Services*. 11(1). p 29-32.

Littlewood B and Strigini L. 1992. The Risks of Software. *Scientific American*. 11. p 38-43.

Ljung L. 1987. System Identification: Theory for the User. Prentice Hall (New Jersey).

Lomas K J, Parand F and Eppel H E. 1989. Definitions of Time in ESP, HTB2 and SERIRES. Applicability Study 1 Research Note 6. Environmental Design Unit. Leicester Polytechnic.

Lomas K J. 1991. Dynamic thermal simulation models of buildings: New method for empirical validation. *Building Services Engineering Research and Technology*. 12(1). p 25-37.

Lomas K J. 1991a. Availability of monitored hourly building performance data for validating dynamic thermal models of buildings. *Building Services Engineering Research and Technology*. 12(2). p 71-74.

Lomas K J, Bloomfield D P, Cole A, Parand F and Pinney A A. 1991. Dynamic Thermal Models: reliability for domestic building design. *Building Services Engineering Research and Technology*. 12(4). p 115-128.

Lomas K J and Eppel H. 1992. Sensitivity analysis techniques for building thermal simulation programs. *Energy and Buildings*. 19. p 21-44.

Lomas K J. 1992. Applicability Study 1: Executive Summary. DeMontfort University report to ETSU S1784.

Lund P D and Peltola S S. 1987. Verification of a CSHPSS simulation program with emphasis on system control. *Solar Energy*. 39(6). p 513-519.

Martin C J and Ruyssevelt P A. 1988. An investigation of heater type and thermostat response on passive solar building performance: Model/data comparison study. ETSU Report S1197-E.

Martin C J and Watson D M J. 1988. Description of the EMC Test Room Facility 1987-88. ETSU Report S1197-P4.

Martin C J. 1991. Detailed Model Comparisons: An Empirical Validation Exercise using SERI-RES. ETSU Report S1197-P9.

Martin C J. 1992. Quantifying Errors in the Predictions of SERI-RES. ETSU Report S1368.

Martin C J and Watson D M J. 1993. Empirical validation of the model SERI-RES using data from test rooms. *Building and Environment*. 28(2). p 175-187.

Martin C J. 1993. A new technique for determining the cause of discrepancies between simulation model predictions and measured data. *Proceedings CIBSE National Conference*. Manchester, UK.

Martin C J. 1993a. Testing thermal models: empirical validation. *BEPAC Information Paper*.

Menicucci D F, Benson M J, Hadfield D W and Otis J S. 1982. An assessment of the accuracy of a home energy analysis system based on a modified TWOZONE computer model. *Building and Environment*. 17(2). p 87-93.

Mihram A G. 1971. Simulation: Statistical Foundations and Methodology. Academic Press, New York.

Mihram G A. 1972. Some practical aspects of the verification and validation of simulation models. *Operational Research Quarterly*. 23(1). p 17-29.

Miller D R. 1974. Sensitivity analysis and validation of simulation models. *Journal of Theoretical Biology*. 48. p 345-360.

Mitalas G P. An experimental check on the weighting factor method of calculating room cooling load. *ASHRAE*.

Morck O C. 1985. Modelling and simulation of solar heating systems. Thermal Insulation Laboratory. Technical University of Denmark. Report no. 170.

Morck O C. (ed.) 1986. Passive and hybrid solar low energy buildings: simulation model validation using test cell data. Thermal Insulation Laboratory. Technical University of Denmark. Report no. 176.

Naylor T H and Finger J M. 1967. Verification of Computer Simulation Models. *Management Science*. 14(2). p 92-101.

Neter J, Wasserman W and Kutner M H. 1990. Applied linear statistical models. Irwin, Illinois.

Ogata K. 1967. State Space Analysis of Control Systems. Prentice Hall Inc.

Onion C T. (ed.) 1973. The Shorter Oxford English Dictionary on Historical Principles. Clarendon Press, Oxford.

Owens D H. 1981. Multivariable and Optimal Systems. Academic Press.

Palomo E and Téllez F M. 1991. Non-parametric tests on the residuals for model validation purposes. IER-CIEMAT.

Papoulis A. 1977. Signal Analysis. McGraw Hill.

PASSYS 1989. Subgroup Model Validation and Development: Final Report. CEC DG XII report 033-89-PASSYS-MVD-FP-017.

Peuportier B. 1989. The simplified simulation approach for the thermal design of multizone buildings: First validation results. University of Stuttgart.

Popper, K R. 1959. *The Logic of Scientific Discovery*. Hutchinson of London.

Press W H, Flannery B P, Teukolsky S A and Vetterling W T. 1989. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press.

Ramdani N. 1994. *Validation Experimentale et Analyse de Signaux: Developpement d'une methodologie de comparaison modele/mesures en thermique du batiment*. PhD thesis at Université Paris XII.

Reckow K H. 1987. Validation of simulation models: Philosophy and Statistical Methods of Confirmation. *Encyclopedia of Systems and Control* edited by: Singh M G. Pergamon Press.

Robinson E A. 1967. *Multichannel Time Series Analysis with Digital Computer Programs*. Holden-Day.

Robinson P and Littler J. Thermal simulation models: Comparison with monitored data for houses with large areas of advanced glazing. *Building Services Engineering Research and Technology*. 13(2). p 61-70.

Sage A P. 1987. Validation. *Encyclopedia of Systems and Control* edited by: Singh M G. Pergamon Press.

Sage A P and Melsa J L. 1971. *System Identification*. Academic Press.

Sargent R G. 1982. Verification and Validation of Simulation Models. *Progress in Modelling and Simulation* edited by: Cellier F E. Academic Press.

Schruben L. 1980. Establishing the Credibility of Simulations. *Simulation*. 34. p 101-105.

SERI-RES Version 1.2. Operating Manual. Solar Energy Research Institute.

Shannon R E. 1975. *Systems Simulation: The Art and Science*. Prentice Hall.

Sluce A J A. 1982. Modelling Reality. *Building Services*. 4(6). p 28-29.

- Sokal R R and Rolf F J. 1969. Biometry. Freeman, San Francisco, California.
- Sorrel F Y, Luckenback T J and Phelps T L. 1985. Validation of hourly building energy models for residential buildings. *ASHRAE Transactions*. 91(2).
- Stoer J and Bulirsch R. 1980. Introduction to Numerical analysis. Springer Verlag.
- Stefanizzi P, Wilson A and Pinney A. 1990. Internal long-wave radiation exchange in buildings: Comparison of calculation methods: I Review of algorithms. *Building Services Engineering Research and Technology*. 11(3). p 81-85.
- Stefanizzi P, Wilson A and Pinney A. 1990. Internal long-wave radiation exchange in buildings: Comparison of calculation methods: II Testing of algorithms. *Building Services Engineering Research and Technology*. 11(3). p 87-96.
- Strachan P and Clarke J A. 1991. Towards an interactive model validation facility. *Proceedings Building Simulation '91*. p 563-569. Nice, France.
- Tasdemiroglu E and Tinaut D. 1985. Comparison of an analytical model with experimental results for a passive solar thermal storage wall. *Solar Energy*. 35(3). p 283-286.
- Thomas P J and Richards R J. 1986. Model Validation. *Transactions of the Institute of Measurement and Control*. 8(4). p 170.
- Uglow C E. 1982. Energy use in dwellings: an exercise to investigate the validity of a simple calculation method. *Building Services Engineering Research and Technology*. 3(1). p 35-39.
- Vemuri V. 1978. Modelling of Complex Systems: An Introduction. Academic Press, London.
- Wai F M, Letherman K M and Burberry P J. 1982. Validation and calibration techniques for hybrid computer simulations in the prediction of seasonal energy consumptions. *Proceedings of CIB Third International Symposium*. Dublin, Ireland.

Waters J R. 1980. The experimental verification of a computerised thermal model for buildings. *Building Services Engineering Research and Technology*. 1(2). p 76-82.

Waters J R and Wright A J. 1985. Criteria for the distribution of nodes in multi-layer walls in finite difference thermal modelling. *Building and Environment*. 20(3). p 151-162.

Watson, D M J. 1984. The use of a shadow band to estimate the direct normal component of solar radiation from global and diffuse measurements. Polytechnic of Central London Research in Building Group report RIB/84/MW/2.

Wetherill G B. 1986. Regression Analysis with Applications. Chapman and Hall Ltd.

Wiltshire T J. 1983. Validation and building energy simulation models. *Sun at Work in Britain*. 16. p 19-22 26-28.

Wortmann D, O'Doherty B and Judkoff R. 1981. Implementation of an analytical verification technique on three building energy analysis codes: SUNCAT 2.4, DOE 2.1 and DEROB 3. Solar Energy Research Institute report CONF-S10405-6.

Wright A J, Bloomfield D and Wiltshire T J. 1992. Building simulation and building representation: Overview of current developments. *Building Services Engineering Research and Technology*. 13(1). p 1-11.

Zaheer-Uddim M, Seth D and Fazio P. 1989. Inter-model comparisons between three PC programs and BLAST. *Energy and Buildings*. 13(3). p 210-216.

Zeigler B P. 1976. Theory of modelling and simulation. Wiley, New York.

Zmeureanu R, Fazio P and Haghighat F. 1987. Analytical and inter-program validation of a building thermal model. *Energy and Buildings*. 10(2). p 121-133.